ESTABLISHING A STANDARDIZED SET OF BASE-LEVEL TRANSPORTATION METRICS

THESIS

Kyle D. Marsh, B.S.
First Lieutenant, USAF

AFIT/GTM/LAR/95S-8

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GTM/LAR/95S-8

ESTABLISHING A STANDARDIZED SET OF BASE-LEVEL TRANSPORTATION METRICS

THESIS

Kyle D. Marsh, B.S.
First Lieutenant, USAF

AFIT/GTM/LAR/95S-8

AFIT/GTM/LAR/95S-8

ESTABLISHING A STANDARDIZED SET OF BASE-LEVEL TRANSPORTATION METRICS

THESIS

Presented to the Faculty of the Graduate School of Logistics

and Acquisition Management of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Transportation Management

Kyle D. Marsh, B.S.
First Lieutenant, USAF

September 1995

Approved for public release; distribution unlimited

Table of Contents

List of Figures

List of Tables

AFIT/GTM/LAR/95S-8

Abstract

Previous research in transportation performance measures found current measures in the operations and maintenance functions inadequate by transportation officers Air Force wide. Past research focused on existing measures and only two transportation functions. Conversely, this research evaluates existing and proposed measures for all base-level transportation functions. The goal of this research is to establish a standardized set of transportation metrics. Metrics are a measuring tool and a component of strategic planning; metrics help organizations meet goals related to customer satisfaction. The research's goal was accomplished by surveying enlisted personnel, officers, and civilians serving in the Air Force. The sample consisted of transportation personnel (transporters) and customers of transportation. The preferences of the subgroups regarding all existing and proposed measures on a survey instrument were compared to establish the set of transportation metrics. The findings indicate a high correlation between transporters and customers; both groups agreed on their preferences of most measures. The findings also indicate a dissatisfaction with existing transportation measures and an eagerness for more meaningful measures.

ESTABLISHING A STANDARDIZED SET OF

BASE-LEVEL TRANSPORTATION

METRICS

I. Introduction

Chapter Overview

The leadership of the United States Air Force recognizes that continuous improvement is a worthy goal. The adoption of Total Quality Management (TQM) in the Air Force resulted in a change in management philosophy and strategy that sought to continuously improve day-to-day activities (processes). With continuous process improvement, all processes, activities, and resources are subject to scrutiny and change. Dorsey J. Talley succinctly stated, "if you cannot measure it, you cannot control it. If you cannot control it, you cannot manage it -- it's as simple as that. Measurement truly separates a successful improvement process from one that fails" (Talley, 1991:xi). To ensure continuous process improvement, a tool is needed to measure the system, policy, or product, so that trends may be observed and appropriate actions taken. This measurement tool is known as a "metric."

Metrics are "measurements made over time, that communicate vital information about the quality of a process, activity, or resource" (AFMC Pamphlet 74-9:11). They are meaningful measures that compel appropriate action. The ability to implement change based on accumulated data distinguishes between a measurement and a metric (AFMC Pamphlet 74-9:28). To be useful, a measure must also be associated with an organization's goals.

To identify and describe the responsibilities of the base transportation function, a functional organizational chart for a typical Transportation Squadron is shown in Figure 1.1. These responsibilities can be subdivided into categories as follows: the traffic management branch moves personnel, personal property, and cargo. The following paragraphs describe the various responsibilities of the traffic management branch:

Personnel Movement. The major transportation activities involved in the movement of personnel are counseling, routing, reservations and travel documentation, air terminal processing, lost and found baggage processing, and management reporting.

Personnel Property Movement. The major activities involved in shipping personnel property are entitlement counseling, inspection, inventory, shipping and receiving, non-temporary storage, and claims and management reporting.

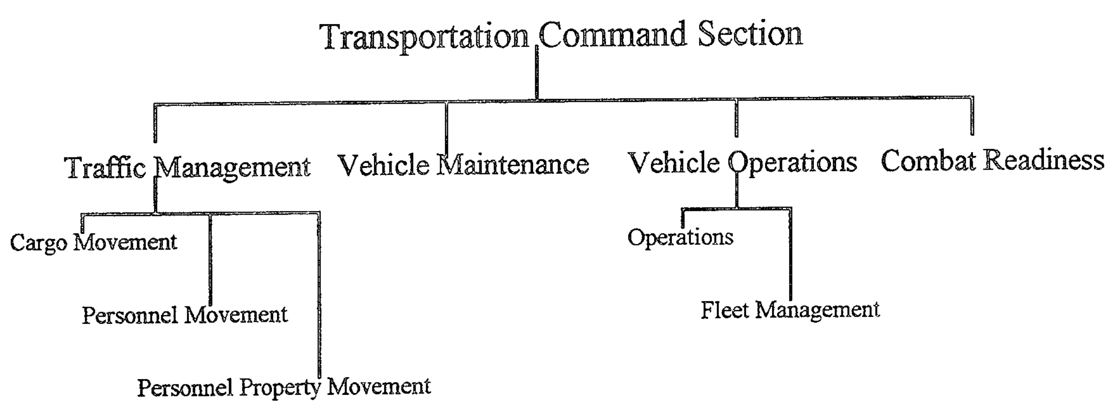

Figure 1-1. Organizational Chart for a Typical Transportation Squadron

Cargo Movement. The major activities involved in the movement of cargo are shipment planning, consolidation and packing, expediting and tracing, terminal processing, and management reporting.

The vehicle operations branch operates the government-owned vehicle fleet for the base. Every base vehicle must be obtained through and accounted for by this office.

Vehicle Operations. The major activities involved in the operation of vehicles includes providing taxi service, licensing military driver's, and dispatching.

Fleet Management. The major activities involved in the management of the vehicle fleet include inspecting vehicles, reporting fleet size and condition, and tracking vehicles. The reports from this branch weigh heavily in the procurement of additional vehicles for an individual base.

The vehicle maintenance branch repairs or modifies government vehicles and vehicle equipment. This branch also maintains maintenance records on all government vehicles, schedules routine (preventive) maintenance, and determines useful life spans.

The combat readiness branch provides training for mobility of equipment and personnel and plays a vital role in any deployment; it inspects, tracks, and transports personnel and equipment from diverse base locations through a processing center and onto aircraft in a timely manner.

The four branches report to a transportation commander. The transportation commander reports the health of the branches to the logistics commander and wing commander in turn. There are hundreds of activities that culminate in an individual branch's performance. Because of the diverse nature of these various activities within each branch, it is difficult to judge performance without effective metrics.

As a component of strategic planning, metrics help managers of organizations meet goals related to customer satisfaction. To meet these goals, metrics measure processes to maximize customer satisfaction. The first step in using metrics involves identifying goals, objectives, and specific tasks. Next, implementing metrics calls for collecting and analyzing data and imposing process changes based on that data.

Background: Performance Measurement

The Industrial Revolution, which started in England at the turn of the eighteenth century and lasted until about 1885, served as the catalyst for the development of the

management discipline (Gray and Smeltzer, 1989: 43). Shortly after this time, Fredrick Winslow Taylor developed Scientific Management, a system of work organization, standardization, and specialization that was designed to measure the efficiency of unskilled labor. His theories involved measuring the most productive worker in a factory and basing a standard on his or her output (Haber, 1964:32). Several decades later, Dr. W. Edward Deming brought Scientific Management up to a new level, and is widely considered the father of modern quality management.

In 1950, Deming left the United States to teach statistical control processes in Japan. It was not until 1978 Deming's reputation first came to the attention of a few auto executives in Detroit who were concerned about losing market share to Japanese automobiles (Evans and Lindsay, 1993:3). Partly due to Deming's advice, the auto industry in the United States has regained some of its market share by building better, fuel-efficient cars. Deming's advice included stressing better, more intelligent ways of doing routine activities. He is well-known for saying, "People [should] work smarter, not harder" (Deming, 1982:3).

In the 1980s, numerous American companies adopted principles of quality set forth by Deming. In 1988, Secretary of Defense Frank C. Carlucci formally adopted Total Quality Management (TQM), a broad concept encompassing many topics that seek to improve management (Bond, 1991:58). This concept resulted in new bureaucracies whose function was to train and implement TQM at every base. In individual transportation squadrons, groups were formed to develop metrics. These metrics, however, were not the measurement tools used in transportation squadrons. Process measurement in Air Force transportation squadrons did not begin in 1988, but in 1968 with the Transportation Integrated Management System (TRIMS).

Background: Measurement in Air Force Transportation

TRIMS was initially intended to create a database capable of interfacing with other functional areas within transportation for on-line data retrieval (Directorate of Transportation, 1968:I-1 - I-5). However, TRIMS was never fully implemented and a subsystem of TRIMS, the Vehicle Integrated Management System (VIMS), was established in its stead in 1970. VIMS data was intended to provide management with a basis to better solve problems (Directorate of Transportation, 1969:10-11). Performance indicators, such as vehicle out-of-commission (VOC) and vehicles down-for-parts (VDM), were established to compare performance of vehicle maintenance units.

The adoption of TQM in the Air Force resulted in a change in management philosophy and strategy that should be linked to performance measures. VIMS measures, such as VOC, do not track performance. An acceptable VOC rating may be achieved by quick-fixes since VOC rates decrease as the time to close work orders decreases. Because TQM suggests that measurement strategy should change, measurement systems should likewise change.

General Issue

In 1989, Lieutenant Kevin N. Brewer conducted a survey which asked whether current measures (VIMS) in vehicle operations and vehicle maintenance flights were useful. His conclusions found that some VIMS measures were perceived as useful, but that further research should be conducted with measures suggested by respondents in the surveys. He concluded that there is a dissatisfaction with VIMS measures, but that transportation commanders were willing to continue using VIMS if no useful alternative was offered.

Brewer asked, how do you measure and quantify customer service in transportation? (Brewer 1989:7) This question called for an analysis of selected measures

based on productivity. His results showed many transporters doubt the usefulness of current measurements based on productivity, and have endorsed an effort to analyze all measurements, existing and proposed measures, to come up with a universal set of useful metrics (Brewer 1989:97).

Brewer's research showed base-level transportation squadrons use performance measures that do not assess productivity accurately. Transportation metrics, quality measures that drive appropriate action, would more accurately assess productivity. Even before TQM was formally adopted, there was an expectation to improve processes and service overall. With ever-increasing attention on quality improvement, accurate metrics are essential.

Specific Problem Statement

In 1968, the Air Force implemented Transportation Integrated Management System (TRIMS). This was the first formalized measurement system for Air Force transportation squadrons. Although every transporter is familiar with the subsequent VIMS measures, their usefulness has not been proven. Brewer's study showed most transportation officers believed that most measures did not convey useful information (Brewer 1989:40).

Since the implementation of TQM in the Air Force in 1988, specific metrics for individual squadrons have been developed. The combination of existing VIMS measures and measures created by individual squadrons in response to TQM has created a bureaucracy in transportation squadrons whose tasks involve measuring and tracking measurements.

A single set of transportation metrics would increase overall efficiency for various reasons. New squadron commanders could immediately recognize what operational goals are, who the customers and processes involved are, and how the processes are measured.

The commander could focus on continuous process improvement based on these data. The focus of the proposed set of metrics would be on the continuous improvement of individual units. The metrics may also be used to compare units, but this alternative is not the main objective. Personnel transferring to another base would benefit from the consistency borne of similar metrics. A standardized set of transportation metrics would eliminate the need for every squadron to commit resources developing their own. Because it would be developed considering Air Force objectives, the set would also direct squadrons toward common goals.

Section Summary. VIMS measures and new measures conceived in response to TQM have not been proven completely useful. Moreover, their continued collection detracts from more important work in transportation squadrons. Consequently, a well-researched singular set of transportation metrics would increase efficiency. The purpose of this research is to analyze past measures and proposed measures to establish a set of ten quality transportation metrics.

Research Questions

The overall objective of this research is to provide a usable set of metrics for Air Force transportation squadrons. This objective may be accomplished by asking the following questions:

1. What measures are currently being used in Air Force transportation squadrons?
2. Upon what qualities should the set of transportation metrics be based?
3. What current transportation measures are useful, and what proposed measures (not currently used) are useful, based on the opinions of transporters and their customers?

The first research question is important because measures are constantly being created as a result of TQM initiatives. Data from VIMS measures are still being collected,

but not every base collects the same measures. The measures suggested by responses to open-ended questions on Brewer's survey offer still more possibilities.

The second question is equally important because measurement qualities differ depending on their intended use. Metrics, as quality tools of TQM, primarily seek to improve processes as their primary focus. This kind of measurement may not be useful for comparing organizations, but it will improve an individual organization. The qualities upon which the set of transportation metrics should be based therefore need to be consistent with a process improvement focus.

The perceptions of transporters and customers of transportation are essential to conclusively determine which measures are useful. Usefulness is determined by accurately measuring productivity and adhering to the quality criteria set forth in investigative question two.

Definition of Performance Measures

Performance may be defined in many ways. The general definition of performance is actual output over expected output. Karlene Crawford gives the following comprehensive definition of a performance measurement indicator:

> [it is] the relative element used to evaluate macro-, micro-, long-term, short-term, flow, static, functional, and overall performance by evaluating the inputs, outputs, transformations (including level of technology), and productivity in a manufacturing or non manufacturing operation. (Crawford 1988:240)

Performance indicators were also researched by Brewer. After examining the opinions of four authors, Brewer determined the following seven criteria comprise a good measure:

1. Validity
2. Coverage
3. Comparability
4. Completeness

5. Usefulness
6. Compatibility
7. Cost Effectiveness

Many similarities can be found when linking Brewer's research to transportation metric development. The criteria for a good performance indicator parallel criteria for a good metric. Productivity and performance definitions comprised the majority of Brewer's research. Metrics, by definition, contain these elements, but also contain quality concepts such as brevity, repeatability, and simplicity.

How Metrics Fit in the Organization

Measurement is an inseparable part of quality improvement. "Making decisive progress without knowing where you were, are, and want to be is nearly impossible" (Fitzgerald, 1989:40). Metrics are vital to successful implementation of quality management.

Metrics are linked to strategic planning. Improvement starts with an organization's mission. Goals are then defined, which are generally related to customer satisfaction; the customer is either the next step in a process or the end-user. Because it is not always possible to elicit continuous feedback from customers, metrics are assumed to incorporate customer satisfaction. That is to say, "if measure X exceeds a predetermined level, then we know the customer is satisfied." In sum, a strategic plan seeks to satisfy customers. Customers are satisfied if a metric exceeds a predetermined level. Therefore metrics are a component of strategic planning.

Scope/Limitations

The results of this research will formulate a response to the overall management question, "Can a useful set of transportation metrics be established that will increase performance and productivity while maintaining the characteristics of a good metric?" This research study will survey transporters and customers of transportation as the target population. Their perceptions of current and proposed performance measures, based on four criteria, will help develop a set of transportation metrics. This research will attempt to evaluate measures for all transportation branches: vehicle maintenance, vehicle operations, traffic management, and combat readiness. Performance measures were gathered from fifteen bases, Headquarters Air Force Material Command, and prior research.

Summary

Quality means continuous process improvement. Metrics are meaningful measures that motivate behavior that will result in continuous process improvement. Measurement is essential to continued progress in the Air Force, including base-level transportation squadrons.

Chapter II examines past and current literature on the subject of measurement in general and metrics specifically. The evolution of measurement in Air Force transportation will also be explored, as well as the origins of metrics within the TQM movement.

Chapter III describes the methodology used to accomplish this study and justifies why this method was chosen. Statistical tests applied to the survey data are defined, explained, and justified.

Chapter IV presents the results and analysis of the study. Tables of important demographic data, percentage descriptive statistics, frequencies of votes for various measures, and results of chosen statistical tests are contained within this chapter.

Chapter V concludes this study and offers specific recommendations. This chapter includes a general outline for subsequent studies of metric use and development in the Air Force

II. Literature Review

Introduction

According to the latest authority on metrics in the Air Force, AFMC Pamphlet 74-9 -- The Metrics Handbook, a metric is defined as:

> A measurement, taken over a period of time, that communicates vital information about a progress or activity. A metric should drive appropriate leadership/management action. Physically, a metric package consists of three parts: (1) An operational definition, (2) Measurement over time, and (3) A presentation package. (AFMC Pamphlet 74-9:13)

In a broad perspective, quality is driven by process improvement, and process improvement is measured by metrics. In this chapter, I will first discuss the origins of measurement as used in the manufacturing industry. I will then describe why measurement is important and what should be measured. Next, I will detail what attributes make a good metric and how metrics are developed. Finally, I will discuss measurement in Air Force transportation, its origins and recent research.

History

Measurement in industry began with Frederick Winslow Taylor's Scientific Management (Grey and Smeltzer 1989: 45). His theories consisted of a system of work organization, standardization, and specialization that was designed to measure the efficiency of unskilled labor. His theories involved measuring the most productive worker in a factory and basing a standard on his or her output (Haber, 1964:32). Though measurement was in its infancy, Taylor posited that a product or service was the result of a series of processes that can be measured, improved, and ultimately perfected. From his ideas, other people found constructive uses for this scientific approach to management.

In the years that followed, measurement was increasingly used to improve manufacturing processes culminating in a revolutionary management philosophy popularized by W. Edward Deming. Deming brought scientific management to a new level and is widely considered the father of modern quality management.

In 1950, Deming left the United States to teach statistical control processes in Japan. In 1978, the reputation of Deming first came to the attention of a few auto executives in Detroit who were concerned about losing market share to Japanese automobiles (Gabor, 1990:31). Deming said, "people [should] work smarter, not harder" (Deming, 1986:1). To Deming, America's quality crisis was symptomatic of a fundamentally outdated management system that focused on short-term results at the expense of the process, the customer, and long-term achievements.

Andrea Gabor analyzed Deming's contributions to American corporations; she suggested that Deming called for "a more systematic approach to pursuing customers and product strategies to replace the mentality of planned obsolescence that worked in the seller's market of the 1950s and 1960s but [was] hobbling American businesses since the 1970s" (Gabor, 1990:7). The historical catalyst for quality in America was the erosion of market share to Japanese corporations.

Deming's contributions enabled American companies to improve quality and regain market share. Facing a competitor's product priced below their cost to build comparable models, Xerox Corporation adopted Deming's techniques in 1979:

> Xerox has undertaken one of the most ambitious quality drives of any major US company. The copier pioneer spent years formulating a corporate quality blueprint that was intended to do nothing less than re-invent [Total Quality Control] by giving it a decidedly American twist. Today Xerox credits that strategy with making it the first US company to recoup lost market share for the Japanese. (Gabor, 1990:188-189)

Prior Research of Base Level Transportation Performance Measures

In 1972, Captains Donald H. Weisert and Sidney H. Clarke conducted research to determine what performance indicators should be used for the base level transportation function. Their thesis was mainly interested in identifying activities of transportation that could be used in a computerized data base. The research attempted to answer what activities should be measured, the quantitative methods for measurement, and what standards should be required for the various activities.

Among their findings was a validation of established Vehicle Integrated Management System (VIMS) measures. An excerpt from the summary section states that the indicators considered useful by the researchers, "percent of vehicles down for parts" and "percent of vehicles down for maintenance," were the same as those considered useful by the majority of the respondents (Weisert and Clarke 1972:66). The research validated existing measurement indicators. Even though the VIMS measures received a mandate from the respondents, this mandate was not permanent. Years later the same population, that is, transportation officers, voiced their dissatisfaction with VIMS measures in another thesis.

In 1989, Lieutenant Kevin N. Brewer conducted a survey which asked whether current measures (VIMS) in vehicle operations and vehicle maintenance flights were useful. He utilized the expertise of 67 transportation officers. His conclusions found that some VIMS measures were perceived as useful, but that further research should be conducted with measures suggested by respondents in the surveys. Brewer admitted his target population (transportation commanders only) and size (67 persons) did not augment the validity of his conclusions. He concluded that there is a dissatisfaction with VIMS measures, but that transportation commanders were willing to continue using VIMS if no useful alternative were offered (Brewer 1989: 97).

A thesis focusing on the VOC rate measure is currently being researched by Captains Larry Audet and Christopher Arzberger. Their preliminary findings show the VOC rate does not reflect the three criteria for a quality performance measurement system: productivity, efficiency, and effectiveness (Arzberger and Audet 1995:14).

Principal Idea: Why Measure?

Deming designed fourteen steps for quality improvement based on six principal ideas. These ideas emphasize the importance of the customer, training, and commitment to quality at all levels. His fourth idea, that "change and improvement must be continuous and all-encompassing" is consistent with his ideas presented in works by contemporary authors (Gabor, 1990:18). In his book, Out of the Crisis, he offers several points to induce transformation of management. His fifth point is to "improve constantly and forever the system of production and service" (Deming, 1986:49). Contemporaries of Deming, Philip Crosby, and J.M. Juran, also note the need for continuous improvement (Tenner and DeToro, 1992:23-24). Measurement is necessary for this process.

Dorsey J. Talley links measurement to process improvement in his book, TQM -- Performance and Cost Measures: The Strategy for Economic Survival. Regarding measurement, he says, "if you cannot measure it, you cannot control it. If you cannot control it, you cannot manage it -- it's as simple as that. Measurement truly separates a successful improvement process from one that fails" (Talley, 1991: xi). We can see the connection between continuous improvement and measurement, but we still need to know what to measure.

Measurement is also important for strategic planning. Robert S. Kaplan and David P. Norton discussed the impact of measures on performance. Talking about present-day managers, they lamented that these managers "rarely think of measurement as

an essential part of their strategy," but later emphasized, "effective measurement, however, must be an integral part of the management process" (Kaplan and Norton, 1993:134). Metrics provide executives with a comprehensive framework that translates a company's strategic objectives into a coherent set of performance measures.

In his book, Planning for Quality, J.M. Juran uses measurement as the key facilitator for strategic decision making. He states, "[measurements] provide an agreed basis for decision making. One purpose of measurement is to provide factual assistance for decision making by diverse human minds" (Juran, 1988:76). Therefore strategic planning can also be aided by metrics.

Measurement is an inseparable part of quality improvement. "Making decisive progress without knowing where you were, are, and want to be is nearly impossible" (Fitzgerald, 1989:40). Quality management techniques have progressed measurably since Taylor's Scientific Management. Metrics are vital to successful implementation of quality management.

Importance of Appropriate Measures

Previously, process improvement has been linked to measurement. It can then be said that the reverse is also true. That is, one method to inhibit process improvement is to measure the wrong elements of a process. Measuring inappropriate elements wastes time, effort, and resources. An organization's goals focus attention on important processes. Tom Peters says we should "measure what's important" (Peters, 1987: 483). Put another way, measuring for the sake of measuring accomplishes little or nothing. In service-oriented industries the activities needing improvement are not always distinct. It is important to specify the particular activity targeted for improvement. In manufacturing, these activities are fairly obvious, but many companies in America and the U.S. Military

are service-oriented. Table 2-1 will show some common metrics in service industries to illustrate what kind of metric is needed in the military.

Table 2-1. Examples of Metrics (Evans and Lindsey 1993:556)

Organization	Sample Metric
Hospital	Lab test accuracy
	On-time delivery of meals and medication
Insurance Company	Claims-processing response time
	Billing accuracy
Post Office	Sorting accuracy
	Percent express mail delivered on time
Ambulance	Response time
Police Department	Incidence of crime in a precinct
	Number of traffic citations
Hotel	Proportion of rooms satisfactorily cleaned
Transportation	Proportion of freight cars correctly routed
	Dollar amount of damage per claim
Auto Service	Percent work completed on time
	Number of parts out of stock

Dennis Kinlaw thinks the goal of measurement is based on three elements. He believes quality measures are those that gauge output, work processes, and customer satisfaction (Kinlaw, 1992:12). Output and work processes are readily understandable, but customer satisfaction requires some elaboration.

According to Juran, customers are "all persons who are impacted by our processes and our products" (Juran, 1988:8). It follows that customers could be anyone within the organization or outside the organization that is impacted in some way by an organization's process.

Deming states "quality is defined by the customer" (Gabor, 1990:18). Because the customer is the most important person in the process, measurements should be tailored to benefit the customer. An example of this kind of measure is, "when you move and use military services, do packers mark each box identifying the content to ensure proper care is taken?" Satisfaction, including courteous treatment and promptness, may be measured in this way. Kinlaw postulates that by measuring the product or service received by the customer, increased satisfaction would result (Kinlaw, 1992:31). Measuring customer satisfaction is difficult and requires the identification of the customer and the customer's need.

Determining what to measure is an important step. Organizations often "borrow" metrics from other organizations to benefit their processes and measure customer satisfaction. Using another organization's measures, however, may not always work. Fortunately, there are established attributes of a good metric that can aid an organization in deciding what to measure.

Aspects of a Good Metric

Many elements make a good metric. The Metrics Handbook suggests eight fundamental attributes:

1. It is accepted as meaningful by the customer.
2. It tells how well organizational goals and objectives are being met through processes and tasks.
3. It is simple, understandable, logical and repeatable.
4. It shows a trend.
5. It is unambiguously defined.
6. Its data is economical to collect.
7. It is timely.

Most importantly:

8. It drives the "appropriate action." (AFMC Pamphlet 74-9:6)

Meaningful to the customer. This attribute simply means that persons or organizations utilizing a metric should find it useful. As discussed earlier, the customer is the most important element in quality management. Because the customer is the most important element, a metric must take into account what is important to the customer.

Organizational goals and objectives. A metric does not exist merely to keep some workers busy. The mission of an organization should be the focus of these measures. By their use, an organization should be able to tell how well it's doing and what areas are deficient. Simple progress toward a goal is not an adequate condition for a metric -- a metric meets organizational goals through continuous process improvement.

Simple, logical, and repeatable. "Simple" means that the data can be understood by anyone. Complex measures are ineffective. Crosby adds, "Quality measurement is only effective when it is done in a manner that produces information people can understand and use" (Crosby, 1979:199). This attribute also means that a metric can be collected by anyone. Complicated measures allow only a few persons to collect the data, and they inhibit widespread use. Their simplicity prevents ownership by a few persons and fosters cooperation. Juran notes, "any such vagueness or complexity becomes a natural source of divisiveness" (Juran, 1988:76). Repeatability adds to validity and helps show a trend over time.

Shows a trend. This element emphasizes that a metric must focus on a process and track it over time. Process variations and deficiencies could then be identified and corrected. Repeated measures show trends and increase accuracy.

Unambiguously defined. Metrics may be used in many ways. One use is for comparative analysis. Because it is important to compare apples to apples and oranges to oranges, metrics must be clearly defined. The words in a metric must infer singular meanings. In an auto shop, the word "error" can be interpreted as a failure to use the proper brand of oil or not tightening some lug nuts. Do these errors carry the same

weight? Juran argues, "error must be defined so that its meaning in the report is unambiguous" (Juran, 1988:78).

Economical to collect. A metric must add value to a process. The benefits gained by using a metric must exceed the effort to collect the data. In discussing the cost-benefit analysis and precision of data, Juran states:

> It is obvious that a balance must be struck between the cost of making evaluations and the value of having them. In part, the application of this criteria relates to the basic question: *Should we measure or not?* More usually the application relates to "precision" of measurement. The unit of measure should be established at that level of precision which enables us to make valid decisions from the data. To go beyond that level of precision usually adds costs without adding value. (Juran, 1988:78)

This attribute leads to the idea that the best way to measure a process is to select a few highly meaningful metrics and discard the rest.

Drives appropriate action. The final attribute of a metric is the most important. As explained earlier, inappropriate action can occur if a metric is poorly defined. If an insurance company measures claims-processing time, as illustrated in Table 2-1, employees may purposely process a claim with little or no investigation to achieve a good metric score. This hypothetical situation is an example of how a metric would impel inappropriate behavior. In an article entitled, "On the Folly of Rewarding A, while Hoping for B," Steven Kerr illustrates the previous scenario with an actual test case. He found as a measure of accuracy in paying claims, an insurance firm measured the percentage of claims paid within two days. This measure made the employees process more claims, including some which should have been disallowed (Kerr, 1975:778). The previous examples show how a measure causes unintended behavior, but what behavior or action should metrics cause?

The Metrics Handbook states that appropriate actions are behaviors displayed by employees that add value to the quality of an organizational product (AFMC Pamphlet 74-9:8). In other words, a metric should lead to, not detract from, an organization's goals.

Section Summary. The various attributes that define metrics clearly differentiate them from other measures. We know that metrics are not simply charts, schedules, snapshots, or counts of an activity. A metric's main defining point is its ability to drive appropriate action. We now know what a metric is, but how is one made?

Metric Development

Past research on metric development has been largely concentrated in manufacturing industries. The development of metrics is so crucial that Hamner and LaFleur dedicated their research primarily to a comprehensive method to develop metrics for service organizations. Their research compared various methods, including the following from the Metrics Handbook:

STEP I. Identify the organization's purpose.
STEP II. Develop the operational definition starting with the organization's customer.
STEP III. Identify and examine the existing measurement systems.
STEP IV. Generate new metrics if existing metrics are inadequate.
STEP V. Rate the metric against the "Eight attributes of a good metric."
STEP VI. Select appropriate measurement tools.
STEP VII. Baseline the process.
STEP VIII. Collect and analyze the data over time.
STEP IX. Finalize the metric presentation.
STEP X. Initiate the process improvement activities.
STEP XI. Continuously monitor and track process improvement. Return to previous steps as appropriate (AFMC Pamphlet 74-9:15).

STEP I. Identify the organization's purpose. This step stresses the importance of first knowing the organization's mission, goals and objectives. The direction of the organization provides a focus that for all measures. This focus should encompass the needs of the organization's customer.

STEP II. Develop the operational definition starting with the organization's customer. It is important to define who, what, when, why, and how of a metric in sufficient detail to permit consistent, repeatable, and valid measurement. The customers' expectations drive the operational definition. This expectation defines the characteristics of the product, service, or process which, if improved, would better satisfy the organization's customers. This is actually an iterative process involving Steps II-VII.

STEP III. Identify and examine the existing measurement systems. Existing measures should be examined prior to establishing new measures to avoid redundancy. Some existing measures may satisfy the organization's requirements already.

STEP IV. Generate new metrics if existing metrics are inadequate. Past research has shown that old measures often are results indicators not tied to an objective or goal (AFMC Pamphlet 74-9:7). These kind of measures are inadequate and should be discarded.

STEP V. Rate the metric against the "Eight attributes of a good metric." Any new metric should satisfy the attributes listed earlier. This step is a checkpoint. The metric developer should continue to Step VI if the metric satisfies the eight attributes, otherwise, he or she should return to Step II and correct the deficiencies.

STEP VI. Select appropriate measurement tools. It is necessary to analyze and display data from metrics so that anyone may understand. Eight tools are listed in the Metrics Handbook, but other statistical and non-statistical tools may also be useful.

STEP VII. Baseline the process. The metric data an organization first gathers serves as a baseline for determining the capability of a process. Later data may be compared against the baseline to determine whether the process has improved or not.

STEP VIII. Collect and analyze the data over time. Collecting metric data should be continuous, and trends should be examined. Trends will tell the manager

which processes are stable, improving, or deteriorating. The data also may show cycles of good and bad performances of an activity which then would suggest a course of action.

STEP IX. **Finalize the metric presentation.** The graphic presentation clearly and concisely communicates how well an organization is performing. The customer may have input deciding how a metric should be displayed, and how detailed he or she wants it.

STEP X. **Initiate the process improvement activities.** This is the most critical step. Process improvement activities should be initiated with the key process owners. It is useful to remember that continuous improvement requires continuous effort.

STEP XI. **Continuously monitor and track process improvement.** Return to previous steps as appropriate. This step is a reminder to assess results of the metric. If a metric becomes obsolete because it no longer contributes to the process improvement, then a new metric may be needed. This situation requires the developer to return to the metric development process.

This method is similar to others in most respects. The key elements consist of defining customers, and developing measures that significantly contribute to an organizations goals. Even so, once a metric is initially developed, there is no guarantee that it is a good metric. A system of measuring metrics is also needed.

Summary

This chapter began with the history of measurement in industry, from the unpolished beginnings of Taylor's scientific management to Deming's quality management, to the countless additions by colleagues and students of Deming. The history of measurement in the Air Force, and specifically transportation was also discussed. The TRIMS and subsequent VIMS measures were discussed, and prior research regarding these measures were shown.

In an effort to provide a background for the standardized set of transportation metrics, the numerous qualities that comprise a good metric were discussed. In this chapter metrics were defined and development of metrics was detailed. These elements provide a fertile ground for the development of a standardized set of metrics for Air Force transportation squadrons.

III. Methodology

Overview

The first two chapters showed that metrics play a vital role assessing and tracking performance of a given activity within any organization, and a common set of metrics for Air Force transportation squadrons would offer many benefits. Given that metrics play a vital role in assessing and tracking performance, this research evaluates VIMS measures and proposed measures to establish a standardized set of Air Force transportation metrics. This chapter addresses the research design executed during the course of this study. This research design furnishes a clear and repeatable investigative effort that produces information designed to answer the investigative questions stated in Chapter I. The origins of measurement in transportation squadrons were reviewed to help justify and answer the research questions. Next, the population for the study was chosen, described, and a plan for sampling this target population was formed. A survey instrument was designed and a plan constructed for collecting and analyzing the data.

Measurement in Air Force Base-Level Transportation Squadrons

Standardized measurement in the Air Force transportation began in 1968 with the Transportation Integrated Management System (TRIMS). In the late 1960s, budget constraints throughout the Air Force called for the effective and efficient utilization of available resources. This concern for better utilization meant managers at all levels needed useful information to operate in the most effective and efficient manner possible. Managers were responsible for broad and complex functions; the complexity made it virtually impossible to know all problems within their area of supervision. This

complexity led to the establishment of an effective management control information system.

In 1968, Headquarters United States Air Force (HQ USAF) initiated an effort to computerize information systems within transportation functions to enhance overall efficiency. This effort was called the Transportation Integrated Management System (TRIMS). TRIMS attempted to create a meaningful data base, standardize the format of output, and interface data with other functional areas (Directorate of Transportation 1968:I1 - I5).

HQ USAF initially intended TRIMS to be a system combining a data base with a program capable of interfacing with other functional areas within transportation for on-line data retrieval (Directorate of Transportation, 1968:I1 - I5). TRIMS was never fully implemented, however, and a subsystems of TRIMS, the Vehicle Integrated Management System (VIMS), was established in its stead in 1970.

VIMS was the first major subsystem completed under the TRIMS program. It was implemented in two phases in mid-1970. Under Phase I, management teams at Air Force Bases in Florida and Texas implemented VIMS on their B-3500 computers. Phase II included all other bases which utilized B-3500 computers (Directorate of Transportation, 1970:1).

VIMS data was intended to provide management with a basis to better solve problems (Directorate of Transportation, 1969:10-11). Performance indicators, such as vehicle out-of-commission (VOC) and vehicles down-for-parts (VDP), were established to compare performance of vehicle maintenance units. Additionally, performance indicators were established to guide management in the analysis of vehicle data. VIMS also included an inquiry capability at all levels of command that enable management to seek data on all categories of motor vehicles by management code (Directorate of Transportation, 1969:11).

The adoption of TQM in the Air Force resulted in a change in management philosophy and strategy that called for more accurate performance measures. VIMS measures, such as vehicle out-of-commission (VOC), do not track performance. VOC rates do not track performance because they are based on the speed of a repair and not the quality of the repair. The assumption that a fast repair is related to the quality of a mechanic's performance is not necessarily true. An acceptable VOC rating may be achieved by quick-fixes since VOC rates decrease as the time to close work orders decreases. Because TQM suggests that measurement strategy should change, there is an implication that measurement systems should likewise change.

There are many VIMS measures, and these measures are discussed later in this chapter. The measures currently used in transportation squadrons are not entirely VIMS measures. Some measures are unique to particular bases and presumably help attain customer satisfaction. These base-specific measures and VIMS measures comprise all current measures, and are discussed later in this chapter. The perceptions of the target population (discussed later) regarding all current and proposed measures comprise the data for this research.

Justification

The format for data collection was a questionnaire. C. William Emory, author of Business Research Methods and expert in research methods, believes questionnaires are best used to measure interests, attitudes, beliefs, feelings, and perceptions of a target population (Emory 1985:158). Additionally, the size of the selected population, and the information needed was large enough to render the use of a telephone survey impractical. According to Don Dillman, the quality of information deteriorates as a telephone survey process lengthens (Dillman 1978:55). Because a telephone survey would be too lengthy for this research, this option was not used.

Research Question One

The purpose of question one, "What measures are currently being used in Air Force transportation squadrons?" was to gather measures upon which to base a comprehensive survey instrument. HQAFMC (LGTX) initiated the process of gathering measures by asking transportation personnel throughout all major commands in the Air Force what transportation measures were used. Three bases, Kirtland, Eglin, and Warner-Robbins, mailed packages to HQAFMC (LGTX) that included charts and graphs of every transportation measures they used. These measures, measures from interviews with personnel at the Air Force Materiel Command Headquarters (HQAFMC), and from personal experiences at Grand Forks AFB, comprised the set of current measures used in this survey. Question one was also answered by gathering measures from prior research on transportation performance measures (VIMS measures), and a thesis by Kevin Brewer in 1989. Brewer's survey posed an open-ended question asking for input from respondents regarding new performance indicators. He specifically asked what measures the respondents thought would enhance efficiency and effectiveness in maintenance and operations functions.

The first measure, "number of vehicles assigned per number of vehicles authorized," was proposed by personnel at Kirtland AFB and was designed to provide information on the actual number of vehicles on a base in relation to what the base should have. The second measure, "number of vehicle accidents with government owned vehicles," gauged the safety record of government vehicle operators who were most often transporters.

The next three measures, "number of vehicle trips supported per number of vehicle trips requested," "percent U-Drive-It requests supported," and "taxi response time (in minutes)" were proposed by personnel at Eglin AFB and evaluated the vehicle operations function.

The following three measures, "number of vehicles in the shop currently undergoing maintenance," "number of vehicles not yet fixed due to lack of proper vehicle parts," and "number of vehicles not yet fixed due to lack of funds to pay for them" are common VIMS measures used by all bases and evaluate the vehicle maintenance function. These measures are also known to transporters as vehicles down-for-maintenance (VDM), vehicles down-for-parts (VDP), and vehicles down-for-funds (VDF) respectively.

Measures nine and ten, "number of vehicles working out of total number vehicles on a base," and "number of vehicles not working out of total number of vehicles on a base," are also common VIMS measures used by all bases and evaluate the vehicle maintenance function. They are also corollaries to one another; that is, if one knows the number of vehicles working, then one also knows the number of vehicles not working by subtraction. The fact that both measures are collected by most bases demonstrates needless repetition. These measures are also known to transporters as vehicles in-commission (VIC) and vehicles out-of-commission (VOC) respectively.

The eleventh measure, "percent of vehicle repairs completed within 24 hours," also evaluates the maintenance function. This measure was proposed by personnel at Warner-Robbins AFB and assumes a high percentage translates into high customer service. However, if repairs are completed hastily to satisfy this measure, shoddy workmanship could result, and the customer would not be satisfied.

The next six measures, "percent of freight (incoming or outgoing) processed within standards," "percent of household goods inspected," "number of inbound and outbound household goods shipments," "number of household goods shipments re-weighed," "percent of household goods shipments delivered on time," and "percent of household goods shipments picked up on time," were proposed in interviews with personnel at HQAFMC and evaluated the traffic management function.

The next three measures, "percent augmentees trained (out of number scheduled)," "tons of mobility items processed," and "percent augmentees trained (out of number of augmentees needed)," were proposed by personnel at Grand Forks AFB and evaluated the combat readiness and resources function.

The following five measures, "average cost per registered vehicle," "percent of vehicles overdue for scheduled maintenance," "number of work orders opened, utilization of administrative vehicles," and "percent of indirect labor," were submitted by personnel at Warner-Robbins AFB and help evaluate the maintenance function.

The twenty-sixth measure, "how well Vehicle Control Officers (VCOs) maintain their fleet," came from personnel at Grand Forks AFB and does not measure any transportation function directly. VCOs are personnel from squadrons other than transportation and oversee the use of government vehicles for their respective units. These personnel act as VCOs as a secondary or tertiary duty to their primary job. VCOs report to and receive training from vehicle operations flights, but do not work for transportation. VCOs performance does reflect the training of the vehicle operations function, but ultimate responsibility rests with the VCO. Therefore, this measure does not reflect any transportation activity directly.

The next two measures, "number of vehicle misuses," and "number of vehicle abuses," were proposed by HQAFMC interviews and were intended to evaluate the performance of the vehicle operations flight. The word "misuse" in the first measure means improperly utilizing a vehicle, like taking a U-Drive-It vehicle to a tavern. "Abuse" means purposely damaging a vehicle, like driving one over unsafe roads.

The last five measures, "percent of equipment available for deployment (in deployable squadrons)," "mean time between failure of vehicles, rejection rate of completed work," "dollars saved from repairing an item rather than replacing it," and "percent of current manpower to required manpower," were proposed in Brewer's thesis.

In his thesis, Kevin Brewer used an open-ended question to solicit new measures from the respondents. In his recommendations, Brewer stated these measures might be useful to transportation squadrons and should be researched further.

Research Question Two

Question two, "Upon what qualities should the set of transportation metrics be based?" was answered by the literature review. The literature review utilized periodicals and books pertaining to quality and AFMC Pamphlet 74-9 to determine critical attributes of an effective metric. A comprehensive review of Air Force regulations, instructions, and other publications provided a basis for the specific transportation qualities. A Defense Technical Information Center (DTIC) literature search ensured the literature review's completeness.

To ensure the survey instrument was understandable, a limited number of qualities were used to help define a good metric. Chapter II detailed eight qualities of a good metric, and the survey instrument utilized a combination of these qualities to form four basic qualities. These four qualities best convey the meaning of a good metric because they incorporate the eight qualities of a good metric. The four criteria and their descriptions follow:

(a) - **It distinguishes health from sickness** -- this means the specific measure of a process signals whether that process is working well or not. A measure of "the percent of workers who smoke" or "the percent of workers under 5 feet 5 inches tall" does not necessarily correlate to poor production; a hundred other factors could affect worker performance.

(b) - **It is simple, understandable, and repeatable** -- this means the measure can be understood by everyone and/or collected by anyone.

(c) - **Workers are able to affect it** -- this means the specific measure can be affected by the persons being measured. An example of a poor measure would be a measure of "how fast the mail is delivered"; this measure can not be affected by the person who takes the letter to the mailbox; the postal service is responsible.

(d) - **It impels appropriate action** -- this means the measure motivates the worker to do the right thing. An example of a poor measure would be "how many pages a secretary types a day"; this measure can not determine whether the secretary is proficient or not. The typewriter may be broken, or the secretary may be typing 100 pages a day, but making errors on 90% of them.

Research Question Three

Question three, "What current transportation measures are useful, and what proposed measures (not currently used) are useful based on the opinions of transporters and their customers?" was answered by gathering data from a representative sample that reflect the perceptions of performance measurements within that population.

The Population and Sample

Before discussing samples and populations, it is necessary to define these terms as statisticians use them. David S Moore and George P. McCabe, authors of Introduction to the Practice of Statistics, define these terms as follows: "The entire group of objects or people about which information is wanted is called a **population**. A **sample** is a part of the population that is actually examined in order to gather information" (Moore and McCabe 1989:278).

To gather appropriate data to evaluate metrics, it is useful to understand the target population who will both use and benefit from this research. The sample of the population was the actual evaluators of the transportation metrics. Because most personnel on any given base use government vehicles, relocate via government-run offices, or practice mobility, they all may be considered part of the statistical population. Persons using transportation services, who are not transporters themselves, are called the "customers" of transportation squadrons. The target population includes transporters and customers of transportation, including all ranks and civilians.

The sample was a cross section of Air Force personnel. Because the statistical sample was all-inclusive, no one group or person was considered inappropriate. Individuals were selected randomly using the personnel records of all military and civilian members in a data base at the Air Force Institute of Technology.

Data from the sampling process were processed manually. The population size consisted of roughly 400,000 military members and 50,000 civilians. According to R.V. Krejcie and D. W. Morgan, authors of "Determining Sample Size for Research Activities," a statistical sample of 300 persons would adequately reflect the opinions of a population of roughly 2000 persons with a 95 percent level of confidence (Krejcie and Morgan 1970:607-610). Krejcie and Morgan further showed for large incremental changes in N (population size), small incremental changes in S (sample population size) are required. For example, a population of one million individuals would require a sample size of 450 persons. Using the formula from Krejcie and Morgan, and a 90 percent level of confidence, a 300 person sample would adequately reflect the opinions of a population of 450,000. The 300 person target population was limited to persons in the continental United States because of mailing time constraints.

Instrument

A preliminary questionnaire was given to ten transportation officers and two civilians familiar both with transportation and metrics. Based on the input from these sources, changes were made to the questionnaire to elicit more useful responses. The survey was designed to answer investigative question three.

This research utilized 300 individuals from two primary subgroups: transporters and customers of transportation. The experience of the respondents included the following Air Force Groups: Logistics, Operations, Support, Communications, and

Medical. Because most individuals use transportation in some form, any person from any base was considered a candidate for this research sample.

The survey consists of two parts. The first section asks for background information of individual respondents. Demographic questions such as gender, age, ethnicity, and rank were not asked. Because the intent of the survey was to establish a single set of transportation metrics, and these classifications did not add to or subtract from the credibility of respondents, it was not useful to classify respondents by gender, age, race or rank. Three categories were deemed useful, however: total years of service, total years in current work center, and primary job experience.

The first two classifications were deemed useful because they infer the degree of respondent's knowledge of measurement in the Air Force and transportation services in general. The third question, "what is your area of primary experience," distinguishes transporters and customers of transportation. This distinguishing factor is useful because those measures transporters think are useful to collect may not be the same as those measures transportation customers think are useful. Because customer satisfaction is a key element of metrics and TQM in general, customer input and opinions are essential.

The second part asks the respondent to judge 33 transportation measures described in the *Question I* section. The respondents' judgments were based on the four qualities described in the *Question II* section. The respondent could select as many of the criteria he or she thought applied to that measure. A fifth option, "it is not a valuable measure" was also provided. Asking respondents to evaluate the measures based on the four quality criteria was meant to facilitate the respondents' decisions regarding the measures. It was assumed that a respondent thinking about the four criteria would choose a measure based on these qualities and not other reasons, like familiarity.

The respondents were then asked to check ten boxes (provided before each measure) for what they thought would constitute the ideal set of transportation metrics in

the Air Force (hereafter, this is referred to as the "voting" portion of the survey). The entire survey appears in Appendix A.

Validation

The test instrument was validated in two ways: expert opinion and a pilot test. These two methods of survey validation are recommended by research experts (Emory, 1983:206). These two approaches contributed to the construct validity of the subject questionnaire. An expert in the field of survey research examined the instrument. Then individuals from the Air Force Materiel Command Headquarters, experts both in metrics and transportation, reviewed the questionnaire. These experts offered advice regarding the wording of the measures and the target population. Because everyone working on an Air Force installation has, in some form, received service from transportation squadrons, these experts thought the target population should include everyone. This idea led to rewording measures so that they could be understood by non-transporters. After numerous revisions the experts believed the questionnaire was valid for assessing respondent's views on transportation metrics.

A pilot test of the revised instrument was accomplished by distributing surveys to ten transportation officers and two civilians familiar with both transportation and metrics. Minor changes were made to the test instrument based on the responses.

Data Collection Plan

The survey instrument was sent to the United States Air Force Survey Control Office, Military Personnel Center, Randolph AFB, Texas for approval. The center approved the questionnaire after requesting minor changes, and a survey control number and expiration date were assigned.

Pre-printed address labels were attached to full size envelopes to facilitate the distribution of the questionnaires. The surveys included a return address on the back

sheet with instructions for respondents to return the surveys to the Air Force Institute of Technology. The questionnaires were mailed in two groups, half on 15 June 1995, and half on 22 June, 1995, from the Operations Office at Building 641 of the Air Force Institute of Technology, Wright-Patterson AFB, Ohio.

Summary

The methodology for this research consisted of a plan to answer the three research questions. Data used to establish question one were researched with a review of relevant literature, past research, and several Air Force bases. Data used for establishing the set of transportation metrics were gathered by a survey of the target population. The data analysis was done by comparing percentages of "yes" votes for each measure. The statistically most important measures for each population were then compared to each other to determine the entire population's preferences. Based on these findings, a universal set of transportation metrics was established.

IV. RESULTS AND ANALYSIS

Chapter Overview

This chapter presents an analysis of the data received from respondents and answers research question three posed in Chapter I of this study. The primary data, or votes, were based on four characteristics established and described in Chapters II and III. The four qualities of a good metric were used to assist and guide respondents; the four qualities were to be used as the sole basis for judging the measures in the survey instrument. The results of the voting are shown in comparative statistics for two subgroups: transporters and customers of transportation. Comparative analysis tests between the subgroups and their respective populations, and between the two subgroups themselves determine the entire populations' preferences of transportation measures. An open-ended question in the survey instrument offered respondents an opportunity to vote for transportation measures they manufactured. Analysis of respondents' measures is intended to assist the establishment of the set of transportation metrics. Before the data is analyzed, however, it is useful to know the sample size and demographics of the respondents.

Sample Population Demographics

The research questionnaires were mailed to 300 individuals throughout the Air Force. Of these, 157 were returned within a month. Even though 181 surveys eventually were returned, the 52% response rate represents the surveys returned early and in a usable form. For the survey to be useful, respondents needed to evaluate all measures and vote for ten measures they thought should be included in the set of transportation metrics. The respondents were split between transporters and customers. There were 73 transporters

and 84 non-transporters. Other demographics establishing the experience base of respondents are presented in Figures 4-1 and 4-2.

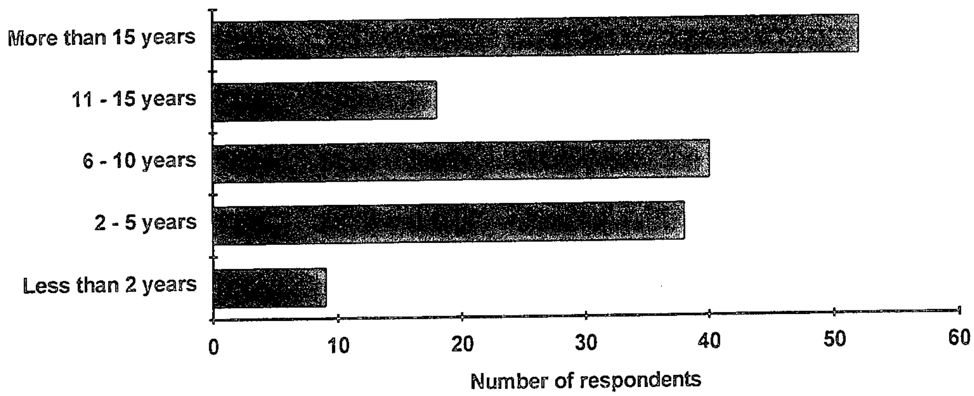

Figure 4-1. Years Worked in the Air Force

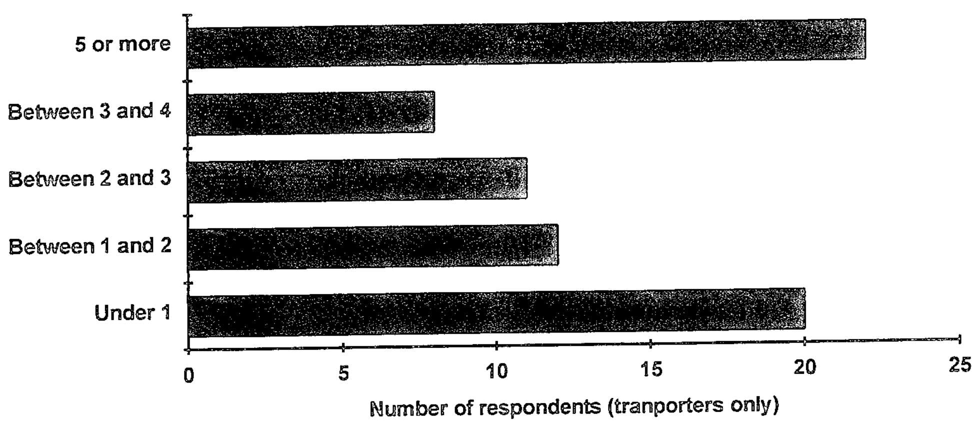

Figure 4-2. Years Worked in Current Work Center (Transporters Only)

Figure 4-1 clearly shows the majority of respondents have many years of experience in the Air Force. Numerous years of experience ensure that most have either moved via a transportation squadron, or have used government transportation services in some way.

Figure 4-2 shows the majority of transporter respondents had more than two years of experience. Only transportation personnel information was used for this Figure. It was assumed a transportation squadron member would be familiar with most transportation measures within a year, but would form a substantiated opinion after two or more years. If the transporter were new to a work center, he or she would probably choose a measure based solely on familiarity. One of the intentions of this research was to ensure respondents chose measures based on a combination of experience and the four criteria of a good metric.

Data Analysis

Data received from this research was analyzed manually. Surveys were divided between transporters and customers of transporters. The votes were then tabulated for each measure. The results for each measure were then divided by the total number of respondents in the respective subgroups to form a proportion. Further analysis compared the preferences of the population of transporters and the population of transportation customers.

Statistical Design

The data consisted of the respondents' ten votes. The votes may be described as simple "yes" or "no" sampling; respondents either voted for a particular measure or they did not. Therefore, this portion is a simple random sample (SRS) (Moore and McCabe 1989:582). The data from the voting portion of the survey were described by percentages of two subgroup sample populations: transporters and customers of transportation. These proportions were then used to calculate an inference about the respective populations (transporters and customers) regarding their preferences of different measures. Confidence intervals about the proportions were calculated to better describe the preferences of the populations as a whole.

$$z = \frac{\hat{p}_1 - \hat{p}_2}{s_p(D)}$$

Where

$$s_p(D) = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

and

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

Variables Defined

$\hat{p}_1$ and $\hat{p}_2$ are the percentages from the samples, $\hat{p}$ is the pooled estimate of the common value of $\hat{p}_1$ and $\hat{p}_2$, z* is the test statistic representing a 95% level of confidence, $s_p(D)$ is the sample distribution, and n_1 and n_2 are the sample population sizes.

This test determined whether the two population proportions were the same. From this test, it was possible to statistically conclude there was agreement between the two populations. This agreement was the basis for deciding which measures were used in the standardized set of transportation metrics.

Comparative Test: Sample of Transporters versus Population of Transporters

Table 4-1 shows the results of a test conducted on each measure to ensure the responses accurately reflected the opinions of the population of transporters as a whole. Because the actual perceptions of the entire population are unknown, a confidence interval was established for each measure. The interval utilized an alpha (α) of .05. In layman's terms, an α of .05 means that there is a 95% confidence that the range or interval established encompasses the actual mean of the population.

Table 4-1. Confidence Intervals of All Measures -- Sample of Transporters versus Population of Transporters

Measure	No. of Votes	Proportion	Confidence Interval (at alpha =.05)
4	41	0.56	(.45, .67)
5	16	0.22	(.13, .31)
6	14	0.19	(.10, .28)
7	37	0.51	(.40, .62)
8	40	0.55	(.44, .66)
9	45	0.62	(.51, .73)
10	19	0.26	(.16, .36)
11	17	0.23	(.14, .32)
12	17	0.23	(.14, .32)
13	15	0.21	(.12, .30)
14	10	0.14	(.06, .22)
15	29	0.4	(.30, .50)
16	5	0.07	(.01, .13)
17	9	0.12	(.05, .19)
18	5	0.07	(.01, .13)
19	39	0.53	(.42, .64)
20	9	0.12	(.05, .19)
21	16	0.22	(.13, .31)
22	4	0.05	(0, .10)
23	50	0.68	(.58, .78)
24	6	0.08	(.02, .14)
25	43	0.59	(.48, .70)
26	6	0.08	(.02, .14)
27	4	0.05	(0, .10)
28	5	0.07	(.01, .13)
29	12	0.16	(.08, .24)
30	15	0.21	(.12, .30)
31	17	0.23	(.14, .32)
32	48	0.66	(.56, .76)
33	18	0.25	(.15, .35)
34	38	0.52	(.41, .63)
35	18	0.25	(.15, .35)
36	42	0.58	(.47, .69)

Table 4-1 suggests that the proportions found by dividing affirmative votes from the sample by the number of transportation respondents accurately reflects the population of United States Air Force transportation personnel. To better understand and read Table 4-1, it is useful to analyze a sample measure.

Measure number 19, from the survey instrument in Appendix A, says, "Percent of household goods shipments picked up on time." From Table 4-1, it can be said that there is a 95% confidence that the population of transporters agree with the sample population of transporters -- the mean proportion of the population of transporters likely falls within 42% and 64%. This interval means that approximately one-half of the population of transporters believes measure number 19, "Percent of household goods shipments picked up on time," is a useful measure and should be included in the set of transportation metrics.

This comparative test correlates the sample of transporters to the population of transporters. It shows that eleven measures, 4, 7, 8, 9, 15, 19, 23, 25, 32, 34, and 36 were statistically preferred by a majority of the population of transporters. A graphic illustration of the transportation votes is shown in Figure 4-3, and the eleven measures appear in Table 4-2.

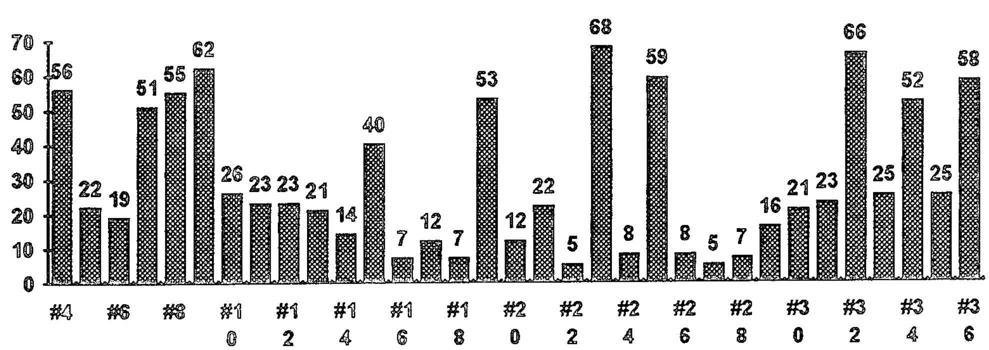

Figure 4-3. Proportion of Transporter Votes

Table 4-2. Measures Preferred by a Majority of Transporters

4. Number of vehicles assigned per number of vehicles authorized.
7. Percent U-Drive-It requests supported.
8. Taxi response time (in minutes).
9. Number of vehicles in the shop currently undergoing maintenance.
15. Percent of freight (incoming or outgoing) processed within standards.
19. Percent of household goods shipments delivered on time.
23. Percent augmentees trained (out of number of augmentees needed).
25. Percent of vehicles overdue for scheduled maintenance.
32. Percent of equipment available for deployment (in deployable squadrons).
34. Rejection rate of completed work.
36. Percent of current manpower to required manpower

Figure 4-3 illustrates graphically the preferences of transporters. Every respondent, as mentioned in Chapter III, was given ten votes to cast among the thirty-three measures plus additional measures provided by the respondent. If respondents had one vote to cast, the proportions would sum to 100 percent. Therefore, the proportions in Figure 4-3 add approximately to 1000 -- 10 votes multiplied by 100 percent less the write-in measures. Transporters compose one half of the target population. To conclude which measures to use in the set of transportation metrics, it is necessary to know the preferences of the customers of transportation.

Comparative Test: Sample of Customers versus Population of Customers

Table 4-3 shows the results of a test conducted on each measure to ensure the responses accurately reflected the opinions of the population of customers of transportation as a whole. As with the populations of transporters, the actual perceptions of the entire population of customers are unknown. A confidence interval with an α of .05 was established for each measure.

Table 4-3. Confidence Intervals of All Measures -- Sample of Customers of Transportation versus Population of Customers

Measure	No. of Votes	Proportion	Confidence Interval (at alpha =.05)
4	26	0.31	(.21, .41)
5	18	0.21	(.13, .29)
6	13	0.15	(.08, .22)
7	14	0.17	(.10, .24)
8	46	0.55	(.44, .66)
9	44	0.52	(.41, .63)
10	52	0.62	(.52, .72)
11	20	0.24	(.15, .33)
12	39	0.46	(.35, .57)
13	10	0.12	(.06, .18)
14	15	0.18	(.10, .26)
15	36	0.43	(.33, .53)
16	5	0.06	(.01, .11)
17	2	0.02	(0, .05)
18	6	0.07	(.02, .12)
19	55	0.65	(.55, .75)
20	18	0.21	(.13, .29)
21	12	0.14	(.07, .21)
22	6	0.07	(.02, .12)
23	22	0.26	(.17, .35)
24	6	0.07	(.02, .12)
25	53	0.63	(.53, .73)
26	6	0.07	(.02, .12)
27	4	0.05	(0, .10)
28	10	0.12	(.06, .18)
29	12	0.14	(.07, .21)
30	15	0.18	(.10, .26)
31	12	0.14	(.07, .21)
32	59	0.7	(.60, .80)
33	46	0.55	(.44, .66)
34	47	0.56	(.45, .67)
35	46	0.55	(.44, .66)
36	52	0.62	(.52, .72)

Table 4-3 suggests that the proportions found by dividing affirmative votes from the sample by the number of customer respondents accurately reflects the population of customers of transportation. The section comparing the sample of transporters to the population of transporters helps to understand and read Table 4-2.

This comparative test correlates the sample of customers of transportation to the population of customers. It shows that twelve measures, 8, 9, 10, 12, 15, 19, 25, 32, 33, 34, 35, and 36 were statistically preferred by a majority of the population of customers. A graphic illustration of the customers' votes is shown in Figure 4-4. A graphic illustration of the customers' votes is shown in Figure 4-4, and the twelve measures appear in Table 4-4.

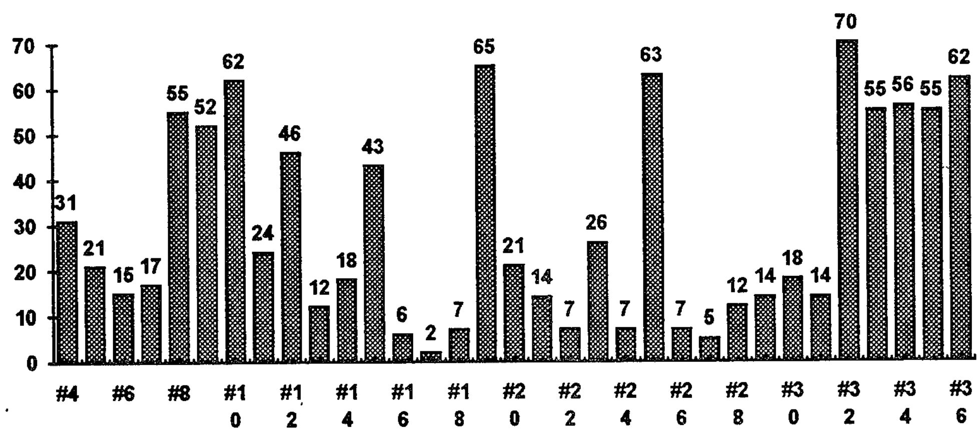

Figure 4-2. Proportion of Customers of Transportation Votes

Table 4-4. Measures Preferred by a Majority of Customers

8. Taxi response time (in minutes).
9. Number of vehicles in the shop currently undergoing maintenance.
10. Number of vehicles not yet fixed due to lack of proper vehicle parts.
12. Number of vehicles working out of total number vehicles on a base.
15. Percent of freight (incoming or outgoing) processed within standards.
19. Percent of household goods shipments delivered on time.
25. Percent of vehicles overdue for scheduled maintenance.
32. Percent of equipment available for deployment (in deployable squadrons).
33. Mean time between failure of vehicles.
34. Rejection rate of completed work.
35. Dollars saved from repairing an item rather than replacing it.
36. Percent of current manpower to required manpower

Comparative Test: Transporters and Customers of Transportation

The goal of this study was to establish a standardized set of transportation metrics. Agreement between the samples of customers and transporters on measures offered in the survey implies that the actual populations agree. The previous two comparative tests showed transporters and customers stated strong preferences for some measures. Additionally, both subgroups preferred a handful of the same measures. A statistical comparative analysis was accomplished for each measure.

The comparison showed statistical agreement between transporters and customers regarding most measures. The null hypothesis for each measure was that the proportions of transporters and customers were equal. Using a two-tailed test and an alpha of .05, the null hypothesis was only rejected when the z statistic exceeded 1.96 or was less than -1.96. Twenty-five measures had calculated z-statistics between -1.96 and 1.96, and therefore were considered statistically equal. On some, however, the statistical tests did not show agreement, that is, the z-statistic was outside the -1.96 and 1.96 range. The proportions of mosts measures were not equal, but close to one another; this statistical

comparison test infers that if the entire populations were surveyed, then the proportions would be equal.

Measures strongly preferred by both populations and statistically agreed upon will be recommended for inclusion in the set of transportation metrics. Measures that indicate strong preferences by one subgroup, but not both, will be analyzed further, and given tentative recommendation for inclusion in the set of transportation metrics.

Appendix B shows the statistical analysis for each measure comparing the preferences of the two subgroups. An example of the statistical analysis appears in Figure 4-3.

Measure #4. Number of vehicles assigned per number of vehicles authorized

<u>Test</u>

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .56 and P_c is the proportion of customers of transportation equal to .31.

H_a: $P_t \neq P_c$

At $\alpha = .05$, using a two-tailed test, we would reject the null hypothesis if $z < -1.96$ or $z > 1.96$

$$z = \frac{.56 - .31}{.08} = 3.1$$

Because 3.1 is greater than 1.96, we reject the null hypothesis in favor of the alternate hypothesis and conclude the two populations are statistically **unequal**.

Figure 4-3. Statistical Comparison Test Example

Analysis

Of the 33 measures, both subgroups had 25 preferences matched by statistical tests. Of these 25, seven measures were preferred by a majority of the populations. The following list comprises these seven measures:

1. The measure of taxi response time.
2. The measure of number of vehicles in the shop currently undergoing maintenance.
3. The measure of percent of household goods shipments delivered on time.

4. The measure of percent of vehicles overdue for scheduled maintenance.
5. The measure of equipment available for deployment (in deployable squadrons).
6. The measure of rejection rate of completed work.
7. The measure of percent of current manpower to required manpower.

Because these seven measures were both agreed upon by both subgroups, and preferred by a majority of the population, they were recommended for inclusion in the set of transportation metrics. The target number of metrics, as stated in Chapter I, was ten. Because the statistical tests did not provide ten measures both agreeable to the subgroups and preferred by a majority of the populations, it was necessary to further analyze the data.

Measures 4, 7, 15, and 23 for transporters and measures 10, 12, 33, and 35 for customers of transportation all were given high preferences by the respective subgroups. These measures are listed on the next page.

Table 4-5. Preferred Measures of the Two Subgroups

Preferred by a majority of transporters 4. The measure of number of vehicles assigned per number of vehicles authorized. 7. The measure of percent of U-Drive-It requests supported. 15. The measure of freight (incoming or outgoing) processed within standards. 23. The measure of percent of augmentees trained (out of augmentees needed). Preferred by a majority of customers of transportation 10. The measure of number of vehicles not yet fixed due to lack of proper vehicle parts. 12. The measure of the number of vehicles not working out of total number of vehicles on a base. 33. The measure of mean time between failure of vehicles. 35. The measure of dollars saved from repairing an item rather than replacing it.

Even though the subgroups did not agree with each other on these measures, the measures all received preferences of over 50% of the respective populations. These high proportions suggested the measures should be included in the set of transportation metrics. To pare this number of measures down to the target of ten metrics, it was useful to look at additional data. Measures offered by respondents in the open-ended questions of the survey instrument reveal similarities to some of the 33 given measures. This similarity, when analyzed, helped determine which ten measures should be used in the set of transportation metrics.

Within the survey instrument, measures 32 through 36 were proposed measures, that is, they are not currently collected and tracked anywhere. The willingness of respondents to vote for these proposed measures indicates a dissatisfaction with existing measures. If respondents were given additional proposed measures, it is likely, based on the evidence just presented, that they would vote for these measures in large numbers as they did for measures 32 through 36 (see Figures 4-1 and 4-2). The additional measures offered by some respondents (Table 4-6) might have received numerous votes if they were included in the survey instrument. Therefore, the additional measures should be weighed heavily in determining the set of transportation metrics.

Additional Measures Offered by Respondents

Respondents were given space for three measures of their own making. Because respondents were not allowed to vote for other respondent's write-in responses, the total number of write-in votes was small (44 total votes). They were further given the opportunity to vote for these measures. In some cases the measure offered was similar to one of the 33 measures. The voting for these "write-in" measures was included, in further analysis, with the data for the measure similar to them in the given 33 measures. This

analysis was done after the tests were conducted to help determine the set of ten transportation metrics.

Table 4-6 lists all the measures the respondents thought would be useful and the measures' respective votes. The first measure in this list, repeat maintenance, indicates the amount of rework for a particular vehicle. From the survey, the measure of mean time between failure of vehicles indicates a similar idea. If a vehicle is not repaired correctly the first time, it would need maintenance again in a short period of time. A low number for the mean time between failures shows either that the vehicle was not fixed properly the first time or that some maintenance was overlooked.

Table 4-6. Write-In Measures and Corresponding Votes

1. Repeat maintenance	7
2. Number of units without required vehicles	1
3. Customer satisfaction	8
4. Mobile Maintenance response time	3
5. Amount of loss and damage	2
6. Pipeline segment time	1
7. Cost per unit of output	3
8. Number of people receiving computer training	1
9. Number of REMS managers with formal training	1
10. Proportion of requirements funded out of those unfunded	1
11. Response time of ordered parts	1
12. Savings from leasing vehicles	1
13. Savings from leasing vehicles	6
14. Percentage of augmentees needed (outside squadron)	3
15. Number of work orders closed per week	1
16. Number of people deployed	4

Either way, the measures, amount of repeat maintenance and mean time between failures, convey similar information; a vehicle requires more maintenance than the customer deems necessary. Because these measures are similar and the majority of one subgroup endorsed

the measure "mean time between failures," this measure should be included in the set of transportation metrics.

Even though one write-in response, "customer satisfaction," received many votes, it cannot be tied to any other measure. The goal of any squadron is customer satisfaction, but it is difficult to measure this concept. Instead, other measures are assumed to ensure customer satisfaction if a certain level is reached.

Measure 14 from Table 4-6, "number of augmentees needed (outside squadron)," is similar to the survey's measure, "percent of augmentees trained (out of number of augmentees needed)." The main difference is that one is an integer and the other is a percentage. Both measures identify the same problem and help track and resolve it. Because this write-in measure mirrors the survey's measure, it too should be included in the set of transportation metrics.

Analysis of the write-in measures shows none of the remaining measures are similar to any measure on the survey. The lack of agreement between customers of transportation and transporters on the measures listed in Table 4-5 and no corroborating measures from Table 4-6 diminishes the desirability of any other measure.

Even though ten measures were sought, the analysis of available data indicates only nine measures were preferred by a majority of the population. Seven were recommended because the majority of both subgroups agreed on them, and two were added to the recommended list because they were essentially identical to measures in the survey instrument which received at least a plurality in one subgroup.

Summary

This chapter analyzed the results of the survey instrument. The demographics of participants showed that most had ample experience with which to draw conclusions. The data from the voting portion was shown in three ways: the transporter subgroup as

correlated to the actual population of all transporters, the customer subgroup as correlated to the entire population of customers of transportation, and the two subgroups correlated against each other. The results of these tests provided a basis for determining the standardized set of ten transportation metrics.

Additional analysis was conducted on the write-in responses. Two write-in measures were similar to measures in the survey instrument. Recommendations were made for measures both attaining a majority and agreement with the subgroups. Two additional recommendations were made for the write-in measures that were both similar to survey measures and engendered at least a majority of one subgroup's support.

V. Conclusions and Recommendations

Introduction

This chapter is comprised of two sections, conclusions and recommendations. The conclusions answer the research questions posed in Chapter I. The second section presents several general recommendations derived from this research effort and recommendations for further research.

Research Questions Answered

Chapter I outlined the course of this research. This chapter introduced metrics and briefly discussed the historical precedent for measurement in business. It also showed the need for metrics in transportation squadrons in the Air Force. Three research questions were posed to guide this research effort. The research questions (listed below for convenience), were explained in Chapter I and answered in subsequent chapters.

Research Questions

1. What measures are currently being used in Air Force transportation squadrons?
2. Upon what qualities should the set of transportation metrics be based?
3. What current transportation measures are useful, and what proposed measures (not currently used) are useful based on the opinions of transporters and their customers?

Chapter II answered research questions 1 and 2. Literature on measurement and metrics was thoroughly reviewed. This information was referred to in conjunction with current measures used in Air Force transportation squadrons. TRIMS and the measures arising out of its subsystem, VIMS, tracked transportation squadrons' performance for 25 years. In a concise form, all the current measures appear in Table 5-1.

Table 5-1. Measures Currently being used in Air Force transportation squadrons

1. Number of vehicles assigned per number of vehicles authorized.
2. Number of vehicle accidents with government owned vehicles.
3. Number of vehicle trips supported per number of vehicle trips requested.
4. Percent U-Drive-It requests supported.
5. Taxi response time (in minutes).
6. Number of vehicles in the shop currently undergoing maintenance.
7. Number of vehicles not yet fixed due to lack of proper vehicle parts.
8. Number of vehicles not yet fixed due to lack of funds to pay for them.
9. Number of vehicles working out of total number vehicles on a base.
10. Number of vehicles not working out of total number of vehicles on a base.
11. Percent of vehicle repairs completed within 24 hours.
12. Percent of freight (incoming or outgoing) processed within standards.
13. Percent of household goods inspected.
14. Number of inbound and outbound household goods shipments.
15. Number of household goods shipments reweighed.
16. Percent of household goods shipments delivered on time.
17. Percent of household goods shipments picked up on time.
18. Percent augmentees trained (out of number scheduled).
19. Tons of mobility items processed.
20. Percent augmentees trained (out of number of augmentees needed).
21. Average cost per registered vehicle.
22. Percent of vehicles overdue for scheduled maintenance.
23. Number of work oders opened.
24. Utilization of administrative vehicles.
25. Percent of indirect labor.
26. How well Vehicle Control Officers (VCOs) maintain their fleet.
27. Number of vehicle misuses.
28. Number of vehicle abuses.

Research question 2 asked which qualities should be used to base the set of transportation metrics. Chapter II detailed the numerous qualities, and Chapter III explained the four used in the survey instrument. The four qualities appear in Table 5-2.

Table 5-2. Qualities Upon Which Metrics Should be Based

1. It distinguishes health from sickness.
2. It is simple, understandable, and repeatable.
3. Workers are able to affect it.
4. It impels appropriate action.

To adequately answer research question 3, a concise methodology was needed. Chapter III provided the methodology for this study. Previous research on transportation performance measures was thoroughly discussed as a basis for Chapter III; a thesis by Kevin Brewer was particularly relevant to this research. Brewer's thesis asked 67 transportation officers to evaluate selected VIMS measures. Brewer's survey instrument, as well as proposed measures from his recommendations, helped form the survey instrument for this research. The additional proposed measures offered from Brewer's thesis appear in Table 5-3.

Table 5-3. Proposed Transportation Measures

1. Percent of equipment available for deployment (in deployable squadrons).
2. Mean time between failure of vehicles.
3. Rejection rate of completed work.
4. Dollars saved from repairing an item rather than replacing it.
5. Percent of current manpower to required manpower (Brewer 1989: 95).

Chapter IV answered research question 3. The original intent of this research was to provide ten metrics comprising a standardized set for all base-level transportation squadrons. Data from Chapter IV led to several conclusions. Transporters and customers of transportation agreed in their assessment of most (76%) measures. The majorities of both subgroups approved of seven measures, and an analysis of write-in responses led to two additional measures. The analysis concluded only nine measures could be recommended based on the available data. The set of transportation metrics appears in Table 5-4.

Table 5-4. The Set of Transportation Metrics

1. The measure of taxi response time.
2. The measure of number of vehicles in the shop currently undergoing maintenance.
3. The measure of percent of household goods shipments delivered on time.
4. The measure of percent of vehicles overdue for scheduled maintenance.
5. The measure of equipment available for deployment (in deployable squadrons).
6. The measure of rejection rate of completed work.
7. The measure of percent of current manpower to required manpower.
8. Mean time between failure of vehicles indicates
9. Percent of augmentees trained (out of number of augmentees needed)

Recommendations

It would be unreasonable to assume this research solves all the measurement problems in transportation squadrons in the Air Force. However, this research does provide a blueprint for other studies endeavoring to build a standardized set of metrics. This section will offer general recommendations and recommendations for future research.

General Recommendations. The nine metrics identified by this research as the standardized set of transportation metrics could be used as a basis to detect trends and set section objectives once standards have been established. After this baselining process, measurements taken in subsequent years should be compared to the initial measurements to detect changes in productivity. Baselining and subsequent comparisons will help managers identify efficient processes and processes needing correction.

These nine metrics could be used in place of existing measures to compare the effectiveness of Air Force transportation squadrons. Currently all bases are judged and compared by VIMS measures. These measures cannot differentiate between an efficient base repairing all vehicle correctly the first time, and inefficient bases reworking maintenance problems constantly; both bases may have an equal vehicle in-commission rate, but customers are not equally satisfied. These nine measures consider customer service and better reflect the effectiveness of transportation squadrons.

Because the intent of the research was to provide metrics based on the four criteria mentioned earlier in this chapter, the respondents needed to understand what these qualities meant. An assumption was made that the brief explanation in the survey instrument was sufficient to convey this vital information. "Impels appropriate action" was misunderstood by numerous respondents; a majority of the respondents did not use this option, and several actually wrote that they didn't understand this criteria. Using a longer explanation would help respondents understand, but the length itself would deter reading the explanation. Using language not requiring explanation in the survey instrument, or using a different data collection technique, such as the Delphi technique or phone surveys to ensure the meanings of the criteria are understood, would also help respondents understand.

Because the desired population was all military members and civilians of the Air Force, a very large sample was required. An assumption was made that of the 300 surveys mailed, a high proportion would be returned guaranteeing validity of the results. This was an erroneous assumption. The problem would have been abated with more surveys. In a research project using survey instruments for data collection, the number of surveys mailed out should equal at least double the desired sample size.

<u>Recommendations for Future Research</u>. The set of transportation metrics posed in this research is untested. Future research should compare the productivity of several bases using the set of transportation metrics and several bases using VIMS measures. The results of this study would validate the set of transportation metrics and ensure wide acceptability in the Air Force.

The set of transportation metrics applies to base-level transportation squadrons. Aerial port squadrons have different objectives and hence different measures are

appropriate. Another study could develop a standardized set of aerial port transportation metrics.

Air Force squadrons should develop their own standardized set of metrics. This development can be done by a study analyzing the effectivenes of existing measures in other squadrons. Effectiveness may be analyzed in many ways; perceptions of squadrons members and the customers they serve would sufficiently judge current measures. Using the metric development outlined in Chapter II, new measures could be proposed. Evaluation of the proposed measures by squadron members and their customers would foster wider acceptance. The simplicity arising from this action would not only benefit them, but would also benefit regional commanders comparing different squadrons based on customer service.

Appendix A : Survey Instrument

AFMC-SPONSORED STUDY
ON
TRANSPORTATION PERFORMANCE MEASURES

RESPONSE FORM
for
ALL PERSONNEL

INFORMATION ABOUT THIS RESEARCH STUDY

Thank you for agreeing to participate in this research project. Your participation in this survey is strictly VOLUNTARY. Not only will your work experience make an important contribution to the goals of this research project, but it will also provide valuable information in Air Force-wide efforts to improve the service you receive from base-level transportation organizations.

Past research in this area: A previous survey asked whether current measures in vehicle operations and vehicle maintenance flights were useful. Some were found to be useful, and are included in this survey. An open-ended question on that survey asked for input from respondents regarding new performance indicators. The questions in this survey are based on the results and suggestions from the open-ended questions of the previous survey.

Description of the study: The goal of this study is to establish a set of 15 (fifteen) performance measures to be used throughout all transportation squadrons. This will be done by determining what current transportation measures are useful, and what proposed measures (not currently used) would be useful. Some examples of common transportation measures are: percent of dispatches supported, vehicle-in-commission rate, vehicles deadlined for parts (VDP), and percent of household goods shipped on time.

How your responses will be used: The information you provide will help to determine which current measures can be eliminated and/or changed, and which ones can be continued. Furthermore, your responses will help to determine whether these measures will be used in the future. In the long run, it may help the Air Force do a better job improving processes rather than gathering meaningless data.

Confidentiality of your responses: This information is being collected for research purposes only. No one in your unit, base, or MAJCOM will see your responses.

BACKGROUND INFORMATION

Please answer the following questions about your background and job experience. This information will be used to develop a profile of the participants in this study. Your responses will be kept completely confidential. Please record your answers on this survey.

---

1. How many years have you worked for the Air Force?

(a) Less than 2
(b) 2 to 5
(c) 6 to 10
(d) 11 to 15
(e) more than 15

2. About how many <u>years</u> have you worked in your current work center?

(a) Under 1
(b) between 1 and 2
(c) between 2 and 3
(d) between 3 and 4
(e) 5 or more

3. What is your area of primary experience?

(a) Transportation
(b) other Logistics Group organizations (Supply, Maintenance, Contracting, Plans & Programs)
(c) Operations Group
(d) Support Group
(e) Communications Group or Medical Group

You may not be familiar with some of the measures mentioned below. Every measure has something to do with the functions of transportation squadrons, and is worded so that non-transporters may better understand. Please answer each question to the best of your ability.

You have been selected as a user or provider of transportation services. Your opinion, as a customer of transportation squadrons, or a transporter, is valuable to this research. Your input will help reorganize (or entirely re-do) the measures used in transportation squadrons in the Air Force.

For the next 33 questions you will be asked to evaluate each measure. The possible responses include:

(a) - It distinguishes health from sickness -- this means the specific measure of a process signals whether that process is working well or not. A measure of "the percent of workers who smoke" or "the percent of workers under 5'5'' tall" does not necessarily correlate to poor production; a hundred other factors could affect worker performance.

(b) - It is simple, understandable, and repeatable -- this means the measure can be understood by everyone and/or collected by anyone.

(c) - Workers are able to affect it -- this means the specific measure can be affected by the persons being measured. An example of a poor measure would be a measure of "how fast the mail is delivered"; this measure can not be affected by the person who takes the letter to the mailbox; the postal service is responsible.

(d) - It impels appropriate action -- this means the measure motivates the worker to do the right thing. An example of a poor measure would be "how many pages a secretary types a day"; this measure can not determine whether the secretary is proficient or not. The typewriter may be broken, or the secretary may be typing 100 pages a day, but making errors on 90% of them.

(e) - It is not a valuable measure -- this response is self explanatory. Please do not include this response with any other response.

Please mark 10 (ten) of the 33 measures in the boxes ( ) provided before each measure, to indicate which ten YOU think should be included in the standardized set of transportation metrics.

(a) - It distinguishes health from sickness
(b) - It is simple, understandable, and repeatable
(c) - Workers are able to affect it
(d) - It impels appropriate action
(e) - It is not a valuable measure

Read the statement below and select <u>all letters</u> from the above list that apply to the measure.

---

4. The measure of the number of vehicles assigned per number of vehicles authorized is valuable because

(a) (b) (c) (d) (e)

---

5. The measure of number of vehicle accidents with government owned vehicles is valuable because

(a) (b) (c) (d) (e)

---

6. The measure of number of vehicle trips supported per number of vehicle trips requested is valuable because

(a) (b) (c) (d) (e)

---

7. The measure of percent U-Drive-It requests supported is valuable because

(a) (b) (c) (d) (e)

---

8. The measure of taxi response time (in minutes) is valuable because

(a) (b) (c) (d) (e)

---

9. The measure of number of vehicles in the shop currently undergoing maintenance is valuable because

(a) (b) (c) (d) (e)

---

10. The measure of number of vehicles not yet fixed due to lack of proper vehicle parts is valuable because

(a) (b) (c) (d) (e)

---

11. The measure of number of vehicles not yet fixed due to lack of funds to pay for them is valuable because

(a) (b) (c) (d) (e)

(a) - It distinguishes health from sickness
(b) - It is simple, understandable, and repeatable
(c) - Workers are able to affect it
(d) - It impels appropriate action
(e) - It is not a valuable measure

Read the statement below and select <u>all letters</u> from the above list that apply to the measure.

---

12. The measure of the number of vehicles working out of total number vehicles on a base is valuable because

(a) (b) (c) (d) (e)

---

13. The measure of the number of vehicles not working out of total number of vehicles on a base is valuable because

(a) (b) (c) (d) (e)

---

14. The measure of percent of vehicle repairs completed within 24 hours is valuable because

(a) (b) (c) (d) (e)

---

15. The measure of percent of freight (incoming or outgoing) processed within standards is valuable because

(a) (b) (c) (d) (e)

---

16. The measure of percent of household goods inspected is valuable because

(a) (b) (c) (d) (e)

---

17. The measure of number of inbound and outbound household goods shipments is valuable because

(a) (b) (c) (d) (e)

---

18. The measure of number of household goods shipments re-weighed is valuable because

(a) (b) (c) (d) (e)

(a) - It distinguishes health from sickness
(b) - It is simple, understandable, and repeatable
(c) - Workers are able to affect it
(d) - It impels appropriate action
(e) - It is not a valuable measure

Read the statement below and select <u>all letters</u> from the above list that apply to the measure.

---

19. The measure of percent of household goods shipments delivered on time is valuable because

(a) (b) (c) (d) (e)

---

20. The measure of percent of household goods shipments picked up on time is valuable because

(a) (b) (c) (d) (e)

---

21. The measure of percent augmentees trained (out of number scheduled) is valuable because

(a) (b) (c) (d) (e)

---

22. The measure of tons of mobility items processed is valuable because

(a) (b) (c) (d) (e)

---

23. The measure of percent augmentees trained (out of number of augmentees needed) is valuable because

(a) (b) (c) (d) (e)

---

24. The measure of average cost per registered vehicle is valuable because

(a) (b) (c) (d) (e)

---

25. The measure of percent of vehicles overdue for scheduled maintenance is valuable because

(a) (b) (c) (d) (e)

---

26. The measure of number of work orders opened is valuable because

(a) (b) (c) (d) (e)

(a) - It distinguishes health from sickness
(b) - It is simple, understandable, and repeatable
(c) - Workers are able to affect it
(d) - It impels appropriate action
(e) - It is not a valuable measure

Read the statement below and select <u>all letters</u> from the above list that apply to the measure.

---

27. The measure of utilization of administrative vehicles is valuable because

(a) (b) (c) (d) (e)

For question 28, "Indirect Labor" refers to all work that does not include 'turning a wrench' or 'tightening a bolt'. Anything that does not directly help repair a vehicle (administrative duties, training, or going to a Commander's Call) is "indirect labor."

---

28. The measure of percent of indirect labor is valuable because

(a) (b) (c) (d) (e)

---

29. The measure of how well Vehicle Control Officers (VCOs) maintain their fleet is valuable because

(a) (b) (c) (d) (e)

For questions 30 and 31, "misuse" means improperly utilizing a vehicle, like taking a U-Drive-It vehicle to a bar. "Abuse" means purposely damaging a vehicle, like driving one over unsafe roads.

---

30. The measure of number of vehicle misuses is valuable because

(a) (b) (c) (d) (e)

---

31. The measure of number of vehicle abuses is valuable because

(a) (b) (c) (d) (e)

---

32. The measure of percent of equipment available for deployment (in deployable squadrons) is valuable because

(a) (b) (c) (d) (e)

---

33. The measure of mean time between failure of vehicles is valuable because

(a) (b) (c) (d) (e)

(a) - It distinguishes health from sickness
(b) - It is simple, understandable, and repeatable
(c) - Workers are able to affect it
(d) - It impels appropriate action
(e) - It is not a valuable measure

Read the statement below and select <u>all letters</u> from the above list that apply to the measure.

---

34. The measure of rejection rate of completed work is valuable because

(a) (b) (c) (d) (e)

---

35. The measure of dollars saved from repairing an item rather than replacing it is valuable because

(a) (b) (c) (d) (e)

---

36. The measure of percent of current manpower to required manpower is valuable because

(a) (b) (c) (d) (e)

Is there a transportation measure YOU think should be included that does not appear above? Please write any transportation measure YOU think should be included in the set of transportation measures in the four spaces below.

37. The measure of

____

38. The measure of

____

39. The measure of

____

40. The measure of

____

Thank you for taking the time to fill out the evaluation portion of this survey.

Please mark 10 (ten) of the 33 measures in the boxes -- ( ) provided before each measure, to indicate which ten YOU think should be included in the standardized set of transportation metrics.

Thank you again for filling out all portions of this survey. The return address appears on the back page of this survey, so all you need to do is fold it over, staple it, and return it through distribution.

Appendix B : Statistical Comparison Tests

Measure #5. Number of vehicle accidents with government owned vehicles

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .22 and

P_c is the proportion of customers of transportation equal to .21.

H_a: $P_t \neq P_c$

At $\alpha = .05$, using a two-tailed test, we would reject the null hypothesis if

$z < -1.96$ or $z > 1.96$

$$z = \frac{.22 - .21}{.066} = .15$$

Because .15 is not greater than 1.96, we do not reject the null hypothesis and conclude the two populations are statistically equal.

Measure #6. Number of vehicle trips supported per number of vehicle trips requested

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .19 and

P_c is the proportion of customers of transportation equal to .15.

H_a: $P_t \neq P_c$

At $\alpha = .05$, using a two-tailed test, we would reject the null hypothesis if

$z < -1.96$ or $z > 1.96$

$$z = \frac{.19 - .15}{.059} = .667$$

Because .667 is not greater than 1.96, we do not reject the null hypothesis and conclude the two populations are statistically equal.

Measure #7. Percent of U-Drive-It requests supported

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .51 and

P_c is the proportion of customers of transportation equal to .17.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if $z < -1.96$ or $z > 1.96$

$$z = \frac{.51 - .17}{.07} = 4.54$$

Because 4.54 is greater than 1.96, we reject the null hypothesis in favor of the alternate hypothesis and conclude the two populations are statistically **unequal**.

Measure #8. Taxi response time (in minutes)

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .548 and

P_c is the proportion of customers of transportation equal to .554.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if $z < -1.96$ or $z > 1.96$

$$z = \frac{.548 - .554}{.08} = -.07$$

Because -.075 is not less than -1.96, we do not reject the null hypothesis and conclude the two populations are statistically **equal**.

Measure #9. Number of vehicles in the shop currently undergoing maintenance

<u>Test</u>

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .62 and P_c is the proportion of customers of transportation equal to .52.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if
$z < -1.96$ or $z > 1.96$

$$z = \frac{.62 - .52}{.079} = 1.27$$

Because 1.27 is not greater than 1.96, we do not reject the null hypothesis and conclude the two populations are statistically equal.

Measure #10. Number of vehicles not yet fixed due to lack of proper vehicle parts

<u>Test</u>

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to ..26 and P_c is the proportion of customers of transportation equal to .62.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if
$z < -1.96$ or $z > 1.96$

$$z = \frac{.26 - .62}{.06} = -6$$

Because -6 is less than -1.96, we reject the null hypothesis in favor of the alternate hypothesis and conclude the two populations are statistically unequal.

Measure #11. Number of vehicles not yet fixed due to lack of funds to pay for them

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to ..23 and

P_c is the proportion of customers of transportation equal to .24.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if $z < -1.96$ or $z > 1.96$

$$z = \frac{.23-.24}{.068} = .147$$

Because .147 is not greater than 1.96, we do not reject the null hypothesis and conclude the two populations are statistically **equal**.

Measure #12. Number of vehicles working out of total number of vehicles on a base

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .23 and

P_c is the proportion of customers of transportation equal to .46.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if $z < -1.96$ or $z > 1.96$

$$z = \frac{.23-.46}{.08} = -2.875$$

Because -2.875 is less than -1.96, we reject the null hypothesis in favor of the alternate hypothesis and conclude the two populations are statistically **unequal**.

Measure #13. Number of vehicles not working out of total number of vehicles on a base

Test

$H_0: P_t = P_c$ Where P_t is the proportion of transporters equal to .21 and

P_c is the proportion of customers of transportation equal to .12.

$H_a: P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if

$z < -1.96$ or $z > 1.96$

$$z = \frac{.21 - .12}{.059} = 1.54$$

Because 1.54 is not greater than 1.96, we do not reject the null hypothesis and conclude the two populations are statistically equal.

Measure #14. Percent of vehicle repairs completed within 24 hours

Test

$H_0: P_t = P_c$ Where P_t is the proportion of transporters equal to .14 and

P_c is the proportion of customers of transportation equal to .18.

$H_a: P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if

$z < -1.96$ or $z > 1.96$

$$z = \frac{.14 - .18}{.06} = -.667$$

Because -.667 is not less than -1.96, we do not reject the null hypothesis and conclude the two populations are statistically equal.

Measure #15. Percent of freight (incoming or outgoing) processed within standards

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .4 and

P_c is the proportion of customers of transportation equal to .43.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if

$z < -1.96$ or $z > 1.96$

$$z = \frac{.4 - .43}{.08} = .381$$

Because .381 is not greater than 1.96, we do not reject the null hypothesis and conclude the two populations are statistically **equal**.

Measure #16. Percent of household goods inspected.

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .07 and

P_c is the proportion of customers of transportation equal to .06.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if

$z < -1.96$ or $z > 1.96$

$$z = \frac{.07 - .06}{.039} = .256$$

Because .256 is not greater than 1.96, we do not reject the null hypothesis and conclude the two populations are statistically **equal**.

Measure #17. Number of inbound and outbound household goods shipments

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .12 and

P_c is the proportion of customers of transportation equal to

.02.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if
$z < -1.96$ or $z > 1.96$

$$z = \frac{.12 - .02}{.04} = 2.$$

Because 2.5 is greater than 1.96, we reject the null hypothesis in favor of the alternate hypothesis and conclude the two populations are statistically **unequal**.

Measure #18. Number of household goods shipments re-weighed

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .069 and

P_c is the proportion of customers of transportation equal to

.07.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if
$z < -1.96$ or $z > 1.96$

$$z = \frac{.069 - .07}{.04} = -.07$$

Because -.07 is not less than -1.96, we do not reject the null hypothesis and conclude the two populations are statistically **equal**.

Measure #19. Percent of household shipments delivered on time

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .53 and

P_c is the proportion of customers of transportation equal to .65.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if
$z < -1.96$ or $z > 1.96$

$$z = \frac{.53-.65}{.08} = -1.5$$

Because -1.5 is not less than -1.96, we do not reject the null hypothesis and conclude the two populations are statistically **equal**.

Measure #20. Percent of household goods shipments picked up on time

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .12 and

P_c is the proportion of customers of transportation equal to .21.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if
$z < -1.96$ or $z > 1.96$

$$z = \frac{.12-.21}{.06} = -1.5$$

Because -1.5 is not less than -1.96, we do not reject the null hypothesis and conclude the two populations are statistically **equal**.

Measure #21. Percent of augmentees trained (out of number scheduled)

<u>Test</u>

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .22 and

P_c is the proportion of customers of transportation equal to .14.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if

$$z < -1.96 \text{ or } z > 1.96$$

$$z = \frac{.22 - .14}{.061} = 1.3$$

Because 1.3 is not greater than 1.96, we do not reject the null hypothesis and conclude the two populations are statistically equal.

Measure #22. Tons of mobility items processed

<u>Test</u>

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .05 and

P_c is the proportion of customers of transportation equal to .07.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if

$$z < -1.96 \text{ or } z > 1.96$$

$$z = \frac{.05 - .07}{.039} = -.51$$

Because -.513 is not less than -1.96, we do not reject the null hypothesis and conclude the two populations are statistically equal.

Measure #23. Percent of augmentees trained (out of augmentees needed)

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .68 and

P_c is the proportion of customers of transportation equal to .26.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if
$z < -1.96$ or $z > 1.96$

$$z = \frac{.68 - .26}{.08} = 5.25$$

Because 5.25 is greater than 1.96, we reject the null hypothesis in favor of the alternate hypothesis and conclude the two populations are statistically **unequal**.

Measure #24. Average cost per registered vehicle

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .08 and

P_c is the proportion of customers of transportation equal to .07.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if
$z < -1.96$ or $z > 1.96$

$$z = \frac{.08 - .07}{.04} = .25$$

Because .25 is not greater than 1.96, we do not reject the null hypothesis and conclude the two populations are statistically **equal**.

Measure #25. Percent of vehicles overdue for scheduled maintenance

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .59 and P_c is the proportion of customers of transportation equal to .63.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if

$$z < -1.96 \text{ or } z > 1.96$$

$$z = \frac{.59 - .63}{.08} = -.5$$

Because -.5 is not less than -1.96, we do not reject the null hypothesis and conclude the two populations are statistically equal.

Measure #26. Number of work orders opened

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .08 and P_c is the proportion of customers of transportation equal to .07.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if

$$z < -1.96 \text{ or } z > 1.96$$

$$z = \frac{.08 - .07}{.04} = .25$$

Because .25 is not greater than 1.96, we do not reject the null hypothesis and conclude the two populations are statistically equal.

Measure #27. Utilization of administration vehicles

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .055 and

P_c is the proportion of customers of transportation equal to .048.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if
$z < -1.96$ or $z > 1.96$

$$z = \frac{.055-.048}{.035} = .2$$

Because .2 is not greater than 1.96, we do not reject the null hypothesis and conclude the two populations are statistically **equal**.

Measure #28. Percent of indirect labor

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .07 and

P_c is the proportion of customers of transportation equal to .12.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if
$z < -1.96$ or $z > 1.96$

$$z = \frac{.07-.12}{.047} = -1.06$$

Because -1.06 is not less than -1.96, we do not reject the null hypothesis and conclude the two populations are statistically **equal**.

Measure #29. How well Vehicle Control Officers maintain their fleet

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .16 and P_c is the proportion of customers of transportation equal to .14.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if

$$z < -1.96 \text{ or } z > 1.96$$

$$z = \frac{.16-.14}{.06} = .333$$

Because .333 is not greater than 1.96, we do not reject the null hypothesis and conclude the two populations are statistically equal.

Measure #30. Number of vehicle misuses

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .21 and P_c is the proportion of customers of transportation equal to .18.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if

$$z < -1.96 \text{ or } z > 1.96$$

$$z = \frac{.21-.18}{.06} = .5$$

Because .5 is not greater than 1.96, we do not reject the null hypothesis and conclude the two populations are statistically equal.

Measure #31. Number of vehicle abuses

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .23 and

P_c is the proportion of customers of transportation equal to .14.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if $z < -1.96$ or $z > 1.96$

$$z = \frac{.23 - .14}{.062} = 1.45$$

Because 1.45 is not greater than 1.96, we do not reject the null hypothesis and conclude the two populations are statistically **equal**.

Measure #32. Percent of equipment available for deployment (in deployable squadrons)

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .66 and

P_c is the proportion of customers of transportation equal to .70.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if $z < -1.96$ or $z > 1.96$

$$z = \frac{.66 - .70}{.075} = -.537$$

Because -.537 is not less than -1.96, we do not reject the null hypothesis and conclude the two populations are statistically **equal**.

Measure #33. Mean time between failure of vehicles

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .25 and P_c is the proportion of customers of transportation equal to .55.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if $z < -1.96$ or $z > 1.96$

$$z = \frac{.25-.55}{.08} = -3.817$$

Because -3.817 is less than -1.96, we reject the null hypothesis in favor of the alternate hypothesis and conclude the two populations are statistically **unequal**.

Measure #34. Rejection rate of completed work

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .52 and P_c is the proportion of customers of transportation equal to .56.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if $z < -1.96$ or $z > 1.96$

$$z = \frac{.52-.56}{.08} = -.502$$

Because -.502 is not less than -1.96, we do not reject the null hypothesis and conclude the two populations are statistically equal.

Measure #35. Dollars saved from repairing an item rather than replacing it.

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .25 and P_c is the proportion of customers of transportation equal to .55.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if
$z < -1.96$ or $z > 1.96$

$$z = \frac{.25 - .55}{.08} = -3.817$$

Because -3.817 is less than -1.96, we reject the null hypothesis in favor of the alternate hypothesis and conclude the two populations are statistically **unequal**.

Measure #36. Percent of current manpower to required manpower

Test

H_0: $P_t = P_c$ Where P_t is the proportion of transporters equal to .58 and P_c is the proportion of customers of transportation equal to .62.

H_a: $P_t \neq P_c$

At α= .05, using a two-tailed test, we would reject the null hypothesis if
$z < -1.96$ or $z > 1.96$

$$z = \frac{.58 - .62}{.078} = -.5$$

Because -.51 is not less than -1.96, we do not reject the null hypothesis and conclude the two populations are statistically **equal**.

References

Arzberger, Christopher A. and Lawrence P. Audet. Evaluation of the Vehicle Out-Of-Commission Standard for Air Force Vehicle Maintenance Units. MS Thesis, AFIT/GTM/LSM/95S-2. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1995. (AD-A313669).

Bond, David F. "Military Budget Drawdown Spurs Movement Towards Increased Efficiency." Aviation Week & Space Technology: 58-59 (December 1991).

Brewer, Kevin N. Perceptions of Air Force Base-Level Transportation Officers Towards the Effectiveness of Air Force Base-Level Transportation Performance Measurement Indicators. MS Thesis, AFIT/GLM/LSM/89S-3. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1989. (AD-A215617).

Crawford, Karlene M. An Analysis of Performance Measurement Systems in Selected Just-In-Time Operations. Ph.D. Dissertation. University of Georgia, Atlanta GA, 1988. (GA-A12193).

Crosby, Philip B. Quality is Free. New York: McGraw-Hill Book Company, 1979.

Deming, W. Edwards "Improvement of Quality and Productivity Through Action by Management." National Productivity Review: 12-22 (Winter 1981-82).

Deming, W. Edwards. Out of the Crisis. Cambridge MA: Massachusetts Institute of Technology Center for Advanced Engineering Study, 1986.

Department of the Air Force Quality and Reliability Assurance -- The Metrics Handbook. AFMC Pamphlet 74-9, HQ Air Force Material Command, Wright-Patterson AFB OH, 31 August 1993.

Dillman, Don A. Mail and Telephone Surveys. New York: John Wiley and Sons, 1978.

Directorate of Transportation, Headquarters United States Air Force, "Transportation Integrated Management System." Washington DC, August 1968. (mimeographed.)

Directorate of Transportation, Headquarters United States Air Force, "Air Force Recurring Publication 75-1 -- Transportation Brief No. 93," Washington DC, December 31, 1969.

Directorate of Transportation, Headquarters United States Air Force, "Air Force Recurring Publication 75-1 -- Transportation Brief No. 95," Washington DC, June 30, 1970.

Emory, C. Williams. Business Research Methods. Homewood, Ill: Irwin Inc., 1985.

Evans, C. William, and William M. Lindsay The Management and Control of Quality. Minneapolis/St. Paul, West Publishing Company, 1993.

Fitzgerald, R. L. "Implementation of a Total Quality Management Program in Private Industry." A Collection of Technical Papers: AIAA/ADPA/NSIA 1st National Total Quality Management Symposium, Denver CO, November 1-3, 1989.

Gabor, Andrea M. The Man Who Discovered Quality. New York: Times Books, 1990.

Gray, Edmund R. and Larry R. Smeltzer Management the Competitive Edge. New York: Macmillon Publishing Company, 1989.

Haber, Samuel A. Efficiency and Uplift. Chicago IL: University of Chicago Press, 1964.

Hamner, Kenneth L. and Charles A. LaFleur An Exploratory Survey of Methods Used to Develop Measures of Performance. MS Thesis, AFIT/GSM/LAS/93S-6. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1993. (AD-A273945).

Juran J.M Juran on Planning for Quality. New York: The Free Press, 1988.

Kaplan, Robert S., and David P. Norton. "Putting the Balanced Score Card to Work." Quality Progress: 134-142 (September-October 1993).

Kerr, Steven "On the Folly of Regarding A, While Hoping for B." Academy of Management Journal: Vol. 18, 769-783 (December 1975.

Kinlaw, Dennis C Continuous Improvement and Measurement for Total Quality. Homewood IL: Pfeifer & Company, 1992.

Krejcie, R.V., and D.W. Morgan "Determining Sample Size for Research Activities." Educational and Psychological Measurement: 607-610 (1970).

Moore, David S. and George P. McCabe Introduction to the Practice of Statistics: 468, 582-593, Wlt Freeman and Company, 1989

Peters, Tom Thriving in Chaos: Handbook for a Management Revolution. New York: Knopf, 1987.

Talley, Dorsey J. Total Quality Management -- Performance and Cost Measures: The Strategy for Economical Survival. Milwaukee: ASQC Quality Press, 1991.

Tenner, Arthur R., and Irving J. DeToro Total Quality Management: Three Steps to Continuous Improvement. Reading MA: Addison-Wesley Publishing Company, Inc., 1992.

Weisert, Donald H., and Sidney H. Clarke Determination of Performance Indicators for the United States Air Force Base-Level Transportation Function. MS Thesis, SLSR-24-72B. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1972. (AD-750847).

Vita

First Lieutenant Kyle D. Marsh is from Gresham, Oregon. He graduated from the University of Arizona in 1992 with a Bachelor of Science degree in General Business. After receiving his commission into the United States Air Force through the Reserve Officers Training Corps, and completing the Transportation Officers Course, Lieutenant Marsh was assigned to the 319th Transportation Squadron, Grand Forks AFB, North Dakota.

During his tour at Grand Forks AFB, Lieutenant Marsh filled a variety of transportation positions, and was the Chief of the Combat Readiness and Resources Flight. Lieutenant Marsh's next assignment will be at the aerial port at Howard AFB, Panama.

Permanent Address: 3725 NW 2nd Circle
Gresham, OR 97030

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE
September 1995

3. REPORT TYPE AND DATES COVERED
Master's Thesis

4. TITLE AND SUBTITLE
ESTABLISHING A STANDARDIZED SET OF BASE-LEVEL TRANSPORTATION METRICS

5. FUNDING NUMBERS

6. AUTHOR(S)
Kyle D. Marsh, First Lieutenant, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology,
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/GTM/LAR/95S-8

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
HQ AFMC/LGTX
Wright-Patterson AFB, OH 45433

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

Previous research in transportation performance measures found current measures in the operations and maintenance functions inadequate by transportation officers Air Force wide. Past research focused on existing measures and only two transportation functions. Conversely, this research evaluates existing and proposed measures for all base-level transportation functions. The goal of this research is to establish a standardized set of transportation metrics. Metrics are a measuring tool and a component of strategic planning; metrics help organizations meet goals related to customer satisfaction. The research's goal was accomplished by surveying enlisted personnel, officers, and civilians serving in the Air Force. The sample consisted of transportation personnel (transporters) and customers of transportation. The findings also indicate a dissatisfaction with existing transportation measures and an eagerness for more meaningful measures.

14. SUBJECT TERMS
Transportation, Metrics, Performance Measures, Instructional Materials, Transportation Management, Management Training, Job Training, Transportation Personnel, Total Quality Management, Quality Air Force, Quality.

15. NUMBER OF PAGES
99

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL